Egil Haga

Correspondences between Music and Body Movement

Faculty of Humanities
University of Oslo
2008

Correspondences between music and body movement

Egil Haga

Ph.D. thesis
Derpartment of Musicology
University of Oslo
2008

Abstract

This thesis explores the way we perceive correspondences between music and body movement. This is done by analysing similarities between music and movement in terms of specific features in music and movement. The analyses have been performed on recordings where participants were asked to make a movement that they thought matched a short excerpt of musical sound, first in a collection of ‘sound-tracings’ where the participants made a movement on a digital tablet, and second, in a collection of ‘free dance movements’ where the participants were asked to move freely to music. The approach to analysis has been developed on the basis of *ecological theory of perception* and *motor theories of perception* that address the perception-movement link. Included in the theoretical framework is also research on *multisensory perception*.

As a point of departure it is assumed that correspondences is perceived on the basis of similar changes in features in music and movement i.e. that correspondence emerges when music and movement co-evolve and change similarly within the same time window. Changes in features are thus analysed on what we call a chunk level in both music and movement.

The analyses of the ‘sound-tracings’ demonstrate how participants are sensitive to correspondences on the basis of changes in pitch, timbre, mode of production (e.g. sustained vs. iterative) and onset density. The analyses of the ‘free dance-movements’ material show that correspondences emerge when overall activation change similarly in music and movement. On more detailed level correspondences are observed in terms of changes in onset density, articulation, speed, as well as how music and movement is subtly performed as ‘pulling’/’stretching’ or ‘hurried’/’jerky’.

These findings suggest that there is a certain consistency in the way we perceive music-movement correspondences. However, the analyses also clearly demonstrate that correspondences emerge flexibly, i.e. that the same musical excerpt may correspond to different variants of movement. It may therefore be difficult to analyse consistency within a traditional quantitative research paradigm and it is proposed that music-movement relations are best examined by detailed qualitative analysis.

Acknowledgements

This thesis on music and movement relationships was funded by the Faculty of Humanities at the University of Oslo.

First, a special thank to my supervisor Professor Rolf Inge Godøy at the Department of Musicology. He has provided encouragement throughout the study and his kind comments and suggestions during the final phase of writing have been invaluable. Through him I was also invited to be a part of the Musical Gesture Project. Collaboration with this group of researchers has offered a wide range of perspectives on music-movement relations.

The observational studies with video-recordings of music and movement are central to the thesis. I wish to thank the people who volunteered to move their bodies in front of the camera, to play the air-piano or other air-instruments as though this was the most natural thing to do, or explore short excerpts of Ligeti music by moving their bodies. Without their participation it would not have been possible to carry out the study the way I wanted.

The collection of video-recordings and sound-tracings was planned and conducted by Alexander Refsum Jensenius, Rolf Inge Godøy and me. The discussions we had during the process of collecting the material were very inspiring and have had a major influence on my own thoughts about music-movement relations and the way they may be approached analytically. Many thanks to Alexander who found the appropriate technical solutions for data recordings and storage, not least the Musical Gestures Toolbox, which he developed as a part of his own doctoral project and which he generously shared with us.

This thesis focuses on the way music is linked to body movement and suggests that this link strongly affects listening to music. In maintaining a restricted focus such as this other perspectives have been omitted, or at least not been given the attention that they might deserve. A very special thank goes to Alicja for her patience and support during the last, rather long phase of the project and most of all for sharing with me what music means to her, thus reminding me that there are many ways of listening to music.

Egil Haga
Oslo, April 2008

Table of contents

Chapter 1. Introduction to a thesis on music-movement correspondences

1.1 The relationship between music and body movement

The overall theme of this thesis is the relationship between music and body movement as this relationship may be observed and experienced when people move their bodies to music. I see the thesis as being part of the field of *music perception.* More specifically, the objective is to describe instances of music and body movements which occur simultaneously and which may be perceived as corresponding, as though they resembled each other. For example, when we see someone moving to music and say that the music and movements go well together, *what are we seeing in the movements* that is similar to *what we are hearing in the music?*

I shall describe the correspondence itself as well as the aspects of music and movement that may contribute to this correspondence. For illustration we may consider the following statement by the composer Earle Brown from the DVD *Merce Cunningham – A Lifetime of Dance*[1]:

> What happened to me was that Merce just sketched out the structure. And then I virtually gave it up – I made my own structure. […] I hate mickey-moused music to dancing. So I just did my own thing. All I did was to use his beginnings and ends. […] Then, I never forget, somewhere in that piece, Merce had this ballet leap, from out of the wings. He came and went all the way across the stage. And it happened that at that point in the music there was a single violin note. Usually, cliché thinking is; "Oh, I gotta support that with a great, big 'oooah' music". And nothing was more striking to me than that leap accompanied by a single violin harmonic. (*Earle Brown, composer*)

This quote is taken from a section of the DVD when composers are commenting on their collaboration with the choreographer and dancer, Merce Cunningham. In general, it seemed that very little had been planned in advance with respect to the relation between music and dance in Cunningham's performances. The composer was given little more than the total duration of the piece, and perhaps a brief sketch of the structure of the choreography. The composer was then left on his own to work the music out independently of the dancers' rehearsing. Thus, it would appear that Cunningham was not trying to control the effects of music-movement relations in advance. One of the composers describes the music as "more a kind of an environment …the air or the weather of the dance"[2]. However, and despite the absence of trying to control the audio-visual relations, another of the composers interviewed describes what he calls "happy accidents" occurring quite often; in fact, "surprisingly often, and wonderful things happened that wouldn't have happened if the artists had planned"[3]. The moment combining the single violin tone with the great

---

[1] Cunningham (2001). A Lifetime of Dance [DVD]: Winstar TV & Video
[2] Comment by composer Meredith Monk
[3] Comment by composer David Behrman

leap, which Earle Brown describes above, is probably one such "happy accident". Out of nowhere a great, forceful leap coincides with a thin, high-pitched violin harmonic, and the relation that emerges catches one's attention and evokes fascination.

These non-planned, accidental audio-visual relations are contrasted with the music composed for cartoons, referred to as a "mickey-moused" relation in which the music imitates aspects of the visual movements on the screen. For example, a running character slowing down to a halt may be accompanied by a deceleration of a similar rate in the music. The heavy tread of a massive body may be illustrated by big, heavy chords in the music, whereas the light steps of a small body may be mimicked musically by using a high-pitched pizzicato melodic pattern. A character running upstairs is commonly "mickey-moused" by using an ascending scale-like melodic contour.

We can imagine other kinds of similarities between music and gesture, and that music and movement converge on the basis of different aspects. For example, the marching of the musicians in a military band corresponds to the successive beats of the music. Likewise, the step pattern of a couple performing a Wiener waltz corresponds to the pattern of contrasting beats in the music, i.e. the alternation of one heavy and two lighter beats that characterises the three-beat metre.

It would seem that not only the step pattern itself matters, but the way the movements are performed also contributes to correspondence. Members of a military band march with firm, decisive steps; they are certainly not tip-toeing. And the couple dancing a Wiener-waltz exhibit a certain flowing character in their movement across the floor. Additionally, and probably closely related to the movement variants just suggested, our perspective concerning music-movement correspondence may be drawn to socio-cultural codes, i.e. that in a certain kind of music we expect certain stylistic features in the movements.

These examples address issues of music-movement relationships which seem to involve the movements following or accompanying the music. Another kind of constellation in which movement and music are closely related may be observed in the way musicians play on their instruments, i.e. in the relation between sound-producing movements and the resulting musical sound. This is a relationship that we probably take for granted, at least with respect to acoustic instruments; when we see the musician perform a big, heavy movement, we expect a correspondingly heavy and forceful tone to emanate from the instrument. This expectance of music-movement correspondence demonstrates a knowledge, a pre-understanding of music-movement relationships, which may also be observed in the way people pretend to play an instrument by making movements such as in 'air-guitar'-playing.

Between the sound-producing and the sound-accompanying we have the movements of a conductor. On the one hand, these correspond to the beat pattern, the phrasing and the flow of the music, and may be similar to dance movements; on the other hand, the conductor's movements help produce the music, or at least guide, shape or facilitate the performance.

1.2. The empirical material of the thesis

In the present thesis the discussion of correspondences will mainly address sound-accompanying movements as found in two kinds of empirical material[4]:

- Video recordings of free dance-movements
- Sound-tracings

The videos of free dance-movements were recorded as part of an observational study. The participants were asked to improvise movements that they thought matched musical excerpts lasting 15 – 20 seconds. Most of the musical examples used in the study could be described as being non-periodic, twentieth-century Western music in the classical tradition. As an example of this, the series of stills in figure 1 shows one of the participants raising her arms slowly to a position above her head as a response to a sustained, lightly dissonant timbre (from György Ligeti's *Ten Pieces for Wind Quintet*, full reference to be given later). The procedure and instructions used in the study will be explained in greater depth in chapters 7 and 8, in which I shall analyse and discuss correspondences in the material collected from the free dance-movements within the framework of an explorative study methodology.

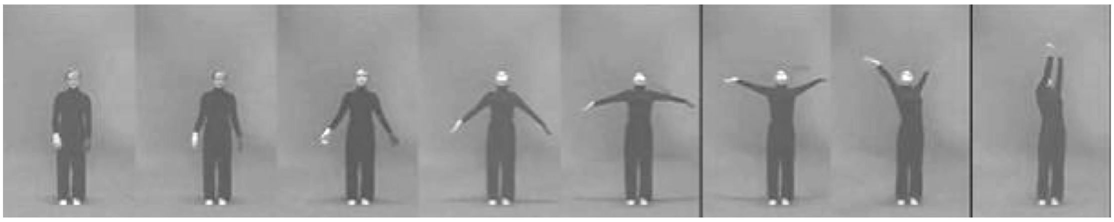


Figure 1. A series of stills illustrating one of the video-recorded movement variants from the collected material referred to as free dance-movements.

The other type of empirical material, the "sound-tracings", was collected in a separate study. Nine participants listened to a series of short sounds (0.5 to 6 seconds), and after each sound they were asked to trace the sound, i.e. to make a movement on a digital drawing tablet that they thought best matched the sound. These responses in the form of drawings were processed and arranged so that the output consisted of a number of displays showing nine different tracings for each sound. Figure 2 shows one such example, i.e. nine interpretations of a trumpet sound which was non-changing in terms of pitch, timbre and loudness:

[4] A collection of examples is available on the CD that was created for this thesis.

Figure 2. Display illustrating a collection of nine different responses to the same trumpet sound as an example of the empirical material referred to as "sound-tracings".

On the basis of this material an inter-rater study was conducted. Twenty people participated in this study. They were shown displays of the type presented above and the corresponding sound was replayed. They were asked to judge to what extent the sound and the different drawings on each display matched or mismatched. The purpose of this study was to describe different kinds of correspondences in the material and to assess whether any kinds of correspondences appeared to be more robust than others. This is the main theme of chapter 6.

The aspects to be addressed in audio-visual analysis later in this thesis are being constrained by the choice of empirical material. As has already been pointed out, I have selected music examples that are characterised by being non-periodic. It should also be noted that I have not chosen music that is constrained by any kind of rhythmical metre, such as the 3/4 metre in a Viennese Waltz or the 4/4 metre in a pop tune, to avoid the music being readily associated with the typical step pattern of any dance style. This implies that I shall not discuss correspondences between a rhythmical "groove" and dance movements as these, for example, may be observed in a pop-music video. In line with this, I shall neither address the phenomenon that have been called *rhythmical entrainment*, i.e. the term *entrainment* understood as a "process whereby two rhythmic processes interact with each other in such a way that they adjust towards and eventually 'lock in' to a common phase and/or periodicity" (Clayton, Sager, & Will, 2004). In musicological research the concept has been used to refer to the human capacity to use the steady beat of a musical rhythm as an external timekeeper and 'fall into' the beat by adjusting a sequence of repeated body movements so that they are synchronised with the beat of the music.

Instead, the thesis will concentrate on movement responses to music that is mainly characterised by textural, timbral and dynamical changes, as well as features that I shall later refer to as the *overall activation level* and *activation contours*.

1.3. The Musical Gestures Project

The observational studies mentioned above were all conducted as a part of the Musical Gestures Project which was conducted at the Department of Musicology at the University of Oslo and has now been completed[5]. Video recordings of air-playing were also included in the music-movement material. Reports about air-playing and sound-tracings have already been published (Godøy, Haga, & Jensenius, 2006a, 2006b). The collection of materials from air-playing, sound-tracings and free dance-movements has also been used to develop computer-based analytical tools (Jensenius, 2006; Jensenius, Godøy, & Wanderley, 2005). The present thesis builds on these publications.

My own thesis is closely connected to the Musical Gestures Project in that it builds on the same theoretical framework, and shares a fundamental interest in the music-movement relationship as captured by the expression *musical gesture*.

1.4. Defining *gesture* and *movement*

When discussing correspondences between music and body movement, I will use the term *movement* to refer to the visual component. The term *gesture* could have been chosen for the same purpose because it is a well-established term in studies of non-verbal communication and sign language (Kendon, 2004; McNeill, 1992). The understanding of *gesture* offered by this research area as meaningful movements of the body is in accordance with the basic perspective on *movement* in this thesis. *Gesture* has also been used in musicological discourse (Cadoz & Wanderley, 2000; Coker, 1972; Hatten, 2004). The term *gesture* will be used when referring to research into non-verbal communication as well as in relation to the expression *musical gesture*. But for the sake of consistency I have chosen to use *movement* as the core term throughout the thesis.

In some cases *movement* will refer to body movement as a phenomenon, whereas in other cases, e.g. in the analyses of music-movement correspondences, the term refers to a delimited unit of body movement, i.e. a *movement chunk*. This is the core meaning of the term in this thesis, and the term *movement* is understood as follows:

- A body movement
- A physical action unit, a movement process that is demarcated by a relatively clear beginning and end. This kind of action unit will be referred to as a *chunk*.
- A *movement* is further characterised by its trajectory in space and its changes in speed (*kinematics*), as well as the way forces are involved and distributed in time (*dynamics*). It is assumed that the perception of body movement as sequences of successive movement chunks emerges on the basis of the way a movement unfolds in time, i.e. the way it is shaped kinematically and dynamically.
- Furthermore, and closely related to the dynamical shaping of the movement, it is assumed that the perception of a *movement* is constrained by the phrasing of the movement, i.e. that the movement process is characterised by a typical phase

[5] See web: http://musicalgestures.uio.no

pattern: *preparatory phase – goal-point – relaxation phase*. In research into non-verbal communication this pattern is referred to as the movement's *peak structure*.

- The movements recorded in our observational studies were performed on the basis of certain instructions, i.e. the participants were typically asked to perform movements that they felt matched the musical sound. In the sound-tracing study, each drawing is understood as capturing a single movement chunk that is intended to correspond to a single sonic chunk. In the free dance movement study the movement sequences typically consist of a number of successive *movement chunks* that are meant to imitate, be similar to or match, or which are simply "inspired" by the musical excerpt. This means that within the context of this thesis a *movement* is understood as a body movement that is intended to relate to musical sound.

The understanding of the term *movement* outlined here is in accordance with the definition of *gesture* applied by researchers of non-verbal communication and sign language. Within this field, two criteria are commonly pointed out:

- A *gesture* is characterised by being a bearer of meaning, i.e. in it being performed with a communicative intent (Cadoz & Wanderley, 2000; Kendon, 2004)
- A *gesture* is characterised by its well-boundedness (a clear beginning and end) and by a phrase structure that is characterised by the way the movement is centred around a moment of stronger emphasis, referred to as a *peak structure* (Kendon, 1996)

The first criterion above allows a distinction to be made between *gesture* and *instrumental actions* (e.g. opening a door), locomotion (e.g. walking across a street) or postural shifts (e.g. moving the head or the eyes to change one's direction of attention). Clearly, the communicative aspect is essential in a research context in which the relation between speech and simultaneous gestures, what are known as co-verbal gestures (McNeill, 1992), are to be examined. For example, we may imagine a situation which involves a man explaining how he opened the door by describing it in words while making a movement with his hand and arm that visualises aspects of how the door was opened. In this case the full meaning emerges on the basis of the interaction between speech and gesture. The movement may thus be defined as a meaningful, well-formed gesture according to the criteria above (McNeill, 1992). Conversely, if the same man a moment later actually opens a door, the similar movement could be defined as a plain, instrumental action.

This distinction between gestures and instrumental actions appears straightforward but is not clear-cut in all cases. Simple instrumental actions may also be perceived as meaningful; not only for their purely functional purpose, but also for the way they convey emotional content, as observed by Kendon:

> In pouring a wine at table, for example, it is possible for the person pouring the wine to 'merely' pour the wine. But it is possible that all the actions involved – raising the bottle to display it, adjusting the angle for the pour, twisting the bottle at the end of the pour to stop a drop of wine from running down the side of the bottle, moving on to the next guest – may be performed so that they are so elaborated with flourishes that they come to be openly recognized as having an

expressive aspect. As this happens, they may come to take on the qualities of gesture (Kendon, 2004; p 9).

Human-body movement is a kind of movement that we are extremely sensitive to in many different respects. From everyday experience we know, for example, that we are able to recognise a friend from the way he or she walks before we actually see the person's face clearly. And we may intuitively ascertain our friend's mood from the way he or she crosses the street, or from the way he or she opens a door. This means that a movement is not only meaningful as a sign in a sign language or as a co-verbal gesture. Even instrumental actions, locomotion and postural shifts may carry meaning in the way they convey emotions and intentions.

1.5. Defining *musical gesture*

Understanding body movement as a mode of articulation that offers a wide range of meanings is central to this thesis. As noted above, leading gesture researchers make a distinction between instrumental actions and gestures on the basis of whether the movement is a carrier of meaning or not. The question is, what kind of *meaning* are we speaking of? The important point to be made here is that making this distinction on the basis of a criterion of meaning that is restricted to *symbolic* aspects of meaning may in many cases be inappropriate and misleading, and this point is worth noting when we speak of music performance. For example, pressing a key on a keyboard is not only an instrumental action; the movement of the arm, hand and finger, and the way this is coordinated with the rest of the body may articulate meaning by suggesting phrasing, dynamics and emotional content. I shall later refer to such features as *non-symbolic* aspects of meaning.

Similar to the sound-tracings and the video-recorded dance-movements, pressing a key on a keyboard is understood as a *musical gesture*. The expression *musical gesture* is in this thesis understood as an actual or imagined body movement that is related to a musical process; and moreover, that it is a body movement that is a bearer of meaning. The expression refers to sound-producing movements such as hitting, pressing, blowing, bowing and plucking, as well as sound-accompanying movements, such as dancing, marching, foot-tapping and swaying to the music.

Imagery of sound-producing or sound-accompanying movements is also included in the concept of *musical gesture*. This brings me to a central theoretical issue underlying this thesis: imagery of sound-producing movements is thought to play a prominent role in the perception of musical sound. This view builds on *motor theories of perception*. Different theoretical variants of motor theories have been proposed, but they share the notion that perceptual processes are intimately linked to (Berthoz, 2000). The central proposition of such theories, on which my own work is based, is that "perception is constrained by action; it is an internal simulation of action" (Berthoz, 2000; p 9). Applying this understanding to perception of music, this implies that sound, musical as well as everyday sounds, is perceived on the basis of imagery of the way the sound process might be produced by movements. Accordingly, I shall assume that the dancers and the sound-tracers in the observational studies, when attending to the musical excerpts, were responding based on the way they imagined the musical sound was being produced by movements, i.e. that their performance of sound-accompanying gestures is linked to an imagery of sound-producing gestures. Acknowledging the role of motor imagery in perception means that the distinction between the sound-producing and the sound-accompanying categories becomes less clear-cut.

1.6. Theoretical framework: motor theory, multisensory perception and ecological theory

What we may collectively refer to as the *motor theory of perception* will be further explained in chapter 3 within the broader framework provided by research into *multisensory perception* and *ecological theory*. But for now, I will say that the following theoretical notions are central to this thesis:

- Perception is constrained by the relation between the perceiver and his/her environment. This means that the perceptual meaning of a given object emerges on the basis of the interaction between the perceiver and the object in the present, and also depends on the understanding the perceiver has acquired of the object during previous encounters.
- Perception is fundamentally *multisensory*, which means that perception of sound is affected by simultaneous input to other senses.
- Perception is constrained by *ecological knowledge*, i.e. sound-movement relations are perceived on the basis of previous experiences of the way sound and movement tend to appear together.
- Perceptions of different categories of sound, both musical and everyday sounds, utilize the same perceptual processes, i.e. they both use motor imagery as a resource in exploring the sonic event perceptually. This suggests that the movements recorded for this thesis may be based on previous experience of sound-movement relations as they appear in both musical and everyday sonic events.

My review of *ecological theory* and *motor theory* will concentrate on the way these theories have developed over the past 20 – 30 years. However, it seems appropriate to briefly mention the theories of perception that were worked out by the Gestalt psychologists early in the twentieth century. First, Gestalt theory seems to anticipate ecological theory in the way it understands perception as an active process by the perceiver. Perceptual experience is viewed as *emergent*; it is not fixed by the physical properties of single elements but emerges on the basis of the way these elements relate to each other. This notion is captured in the following core principle: “the whole is different from the sum of its parts” (Atkinson, Atkinson, & Hilgard, 1983). By including the perceiver and his/her relation to the environment, this fundamental relativity of perception is taken a step further in the *ecological approach to perception* (Gibson, 1986)[6].

[6] However, see Heft’s discussion of the relation between *ecological theory* and *Gestalt theory* in *Ecological Theory in Context: James Gibson, Roger Barker, and the Legacy of William James' Radical Empiricism* (Heft, 2001). Heft proposes that Gestalt psychology differs markedly from

Second, the terms *gesture* and *movement* are in this thesis linked to the understanding of the chunking of time-dependent phenomena, and furthermore, the analyses of music, movement, as well as compounds of music-movement processes, will focus on the way auditory and visual streams are chunked, i.e. that elements combine and interact so that they "melt together" and form temporal *wholes*. These themes were central to the work of the Gestalt psychologists; indeed, at the core of their interest lay the relation between the *element* and the *whole* (Eichert, Schmidt, & Seifert, 1997), and they formulated principles that attempted to describe this relationship and to some degree predict how elements would interact to form a whole (e.g. the principles of *proximity*, *similarity*, and *continuity*, see Snyder, 2000). In sum, there are two ways in which my own thesis is closely related to Gestalt psychology:

- the overall perspective on perception as *emergent*
- the interest in the relation between *elements* and the *whole*

1.7. Research questions and basic hypotheses

The main research question to be addressed in this thesis is:

- *How do perceived aspects of movement interact with perceived aspects of a simultaneous, co-evolving musical sound, so that the two streams are 'seen-heard' as corresponding?*

As a point of departure I shall assume that a sense of correspondence emerges on the basis of:

- *Synchronicity*, i.e. those auditory and visual events that occur close to each other in time and space tend to be perceived as belonging to the same audio-visual event.
- *Similarities*, i.e. that auditory and visual events that appear to be similar (e.g. with respect to *dynamics* and/or *kinematics*, see below) tend to be perceived as belonging to the same audio-visual event.

These two assumptions may be re-phrased to be combined in one general hypothesis:

- Correspondence between movement and music emerges when the auditory and the visual streams evolve similarly over time.

There is one further hypothesis that I wish to discuss. The material from the sound-tracings and the free dance-movements demonstrates that there is considerable diversity in the responses to the same musical excerpts, but that different movement variants may still be perceived to correspond to the same music. On the other hand,

---

ecological theory in the way the Gestaltists maintain a fundamental dualism between the perceiver and the physical world. In ecological theory, the perceiver stands in a direct and intimate relation to the environment as a result of the way the perceptual system has co-evolved with features of the environment, and as a result of socio-cultural processes (Heft, 2001, pp 232-233).

the variants also appear to be similar in certain aspects, so that it seems that there is a degree of consistency in the way the music offers a movement response.

In light of this initial observation it seems that correspondence is on the one hand *flexible*, i.e. a wide range of gestural variants may correspond to the same music; and, on the other hand, *non-arbitrary*, i.e. certain qualities may contribute to a sense of correspondence more consistently than others. This hypothesis has been re-phrased into a second research question:

- *Is it possible in the empirical material to identify features in music and movement that across most of the audio-visual variants seem to consistently contribute to correspondence so that the features may be understood as* non-arbitrary*?*
  And:
  Is it possible in the empirical material to identify features that in some cases contribute prominently to correspondence, but that in other cases do not contribute to the same degree, so that they may be understood as flexible*?*

1.8. The term *correspondence*

Studying the empirical material in this way emphasises that the analysis of correspondences will be approached broadly. The aim is to describe from different angles how correspondences seem to emerge in the material. *Correspondence* will be used as a general term that refers to a qualitative judgement of the relation between music and gesture in an audio-visual example, where the relation is characterised by the way music and gesture are perceived as being similar. The dictionary proposes "be similar to", "be equal to" and "be in harmony with" to explain the term *correspond*. Music and movement may correspond in different ways, and we may think of alternative terms that might further describe the quality and degree of *correspondence*, such as *imitation*, *mimicry, convergence*, *match*, *amalgamation*, etc. Descriptions of the different shadings of correspondence will for each audio-visual variant be a part of the analytical procedure, which means that *correspondence* will not be defined with clear, exclusive criteria.

In this regard, it should be noted that the term *similarity* refers to phenomena that bear a resemblance to each other[7]. In my context, *similarity* means that music and movement resemble each other in certain aspects, whereas they may differ in others. It does not imply an identical match of two objects or events. This implies that this thesis on music-movement correspondences is basically not a study of *sameness*; it is a study of *similarities*, which may include *differences*. In sum, the qualitative judgement of correspondence includes both similarities and differences.

1.9. Features contributing to correspondence

The question is: when we say that music and movement are similar to each other, what do we mean; *which features in music and movement are we referring to?* The features that will be examined theoretically as well as on the basis of analyses of the empirical material may be summarised as follows:

[7] Websters comprehensive dictionary, 2003. Encyclopedic edition.

- *Activation contour*, i.e. the sense of changes in activity level, or energy, in a musical or a gestural excerpt
- *Kinematics*, i.e. the trajectory and the changes in the speed of a movement
- *Dynamics*, i.e. the forces that cause and constrain a movement
- *Chunking*, i.e. the segmentation of continuous musical sound and/or body movements into temporally demarcated units
- *Points of synchronisation* or *synch-points*, i.e. simultaneous accentuations in music and movement

I would hypothesise that music and gesture may be perceived as similar on the basis of these features. Additionally, it seems necessary to introduce the concept of *rhythm* since this is an important experiential aspect that is associated with both music and movement, hence relevant to the analysis of correspondences. To build on a discussion by Waadeland (Waadeland, 2000), rhythm entails a two-sidedness; on the one hand, an element of order and structure, and on the other hand, an element of expressiveness, i.e.:

- a temporal ordering of contrasts, e.g. as in the contrast between a heavy downbeat and a lighter upbeat;
- a sense of flow, e.g. conveyed in descriptive terms used by musicians, such as swing and groove.

In the following I shall refer to the above listed features, including rhythm, as *non-symbolic* aspects of meaning of music and movement. I would propose, considering the nature of the sound-tracings and the free dance-movements, that these features are particularly relevant when discussing correspondences in the collected material.

In this context *non-symbolic* denotes aspects of meaning that do not readily convey verbal meaning in the same way that a *symbol* does, i.e. a symbol understood as a sign or a form that is "chosen to stand for or represent something else"[8]. I understand music and movements to be human forms of articulation that are meaningful in many different ways, ranging from the non-symbolic to the symbolic, a perspective that will be further discussed in chapter 5.

1.10. Overview of the thesis

The present, and the next chapter which contextualises the thesis within music theory and music perception, and which also further discusses the aims and methodology of the thesis, are introductory chapters. The next three chapters present the theoretical framework:

Chapter 3. In chapter 3 I shall review research into multisensory perception, with a focus on audio-visual interaction. I shall also present the basic notions of ecological theory and motor theories of perception. In the concluding part I discuss how the empirical material may be understood in light of these theories.

[8] Websters comprehensive dictionary, 2003. Encyclopedic edition.

Chapter 4. This chapter is devoted to the concepts of *dynamics* and *kinematics*. I start by explaining the concepts of *effort* and *activation* as they are understood as rhythmical-dynamical aspects in Rudolf Laban and Daniel Stern's theories respectively. In the second part of the chapter I shall define the concepts of dynamics and kinematics as they originate from physical science and the way they have been applied to movement studies. Following this, I shall discuss whether the concepts are useful in musical analysis. This leads to a more overarching discussion about music as motion.

Chapter 5. The main theme of this chapter is how both music and movement are meaningful in many different ways. This leads to a discussion of how music and gesture are meaningful in a *non-symbolic* manner, i.e. how processes of chunking and kinematical/dynamical shaping are linked together thus producing *non-symbolic* aspects of meaning.

In chapters 6, 7 and 8 I present annotations of audio-visual data, the empirical material of the thesis. In chapter 6 I discuss correspondences in the sound-tracing material in terms of dynamical and kinematical aspects. In chapter 7 I form observational categories, i.e. features of music and gesture that on the one hand are related to the *dynamics-kinematics* distinction, and which, on the other hand, are based on the concept of *activation*. These features are implemented in an analytical procedure that in chapter 8 is applied to a detailed analysis of selected audio-visual variants from the free dance-movements material. The concluding chapter 9 sums up the main theoretical and analytical perspectives that are presented in the thesis. The appendix provides texts for instructions used in the sound-tracing and inter-rater studies presented in chapter 6.

Chapter 2. Contexts, methodology and relevance

> To relate these two elements [picture and sound], we find a natural language common to both – movement. *Sergei Eisenstein* in *The Film Sense* (1986; p 70)

2.1. Introduction

Questions of correspondence, or similarity, are much discussed in the field of musical multimedia. Correspondence may be based on different aspects, such as symbolic meaning, as well as on rhythmical patterning and flow. In film-music analysis, discussions of correspondence between media also involve issues concerning the relationship between music and images, i.e. whether the relationship is characterised as hierarchical, dialectic, and/or complementary. In my view, the question of similarity between music and images constitutes a point of reference for any discussion of musical multimedia. This view provides a framework of understanding within which my own thesis on music-movement relations is situated.

This thesis is also affiliated to the branch of music theory that is concerned with descriptions of the way individual tones interact to work as music; for example, the way melodic lines and dissonances are treated to produce a specific musical style (Jeppesen, 1992), or the way harmony and instrumentation work together to produce orchestral colours and effects (Rimsky-Korsakoff, 1964). The relation between music and images in film has been approached in a similar manner, for example by Sergei Eisenstein in *The Film Sense* (Eisenstein, 1986), by Theodor Adorno/Hanns Eisler in *Composing for The Films* (Adorno & Eisler, 1994), as well as in a more recent contribution from Randall Meyers in his book *Film Music. Fundamentals of the Language* (Meyers, 1994). These accounts discuss quite specifically the way different musical elements may (or in some cases should) be combined with images so that a desirable result emerges. They also include discussions about the music-image relationship on a more principle level. Less detailed with respect to specific compositional elements are two more recent contributions, Nicholas Cook's *Analysing Musical Multimedia* (Cook, 1998) and Michel Chion's *Audio-Vision. Sound on Screen* (Chion, 1994); however, they both provide in-depth discussions of music/sound-image interactions on a theoretical level as well as with reference to concrete audio-visual analyses.

Considering the main research question of my thesis, *how do aspects of music and movement work together so that they are perceived as corresponding*, I would position the present thesis within the context of music theory, as well as within the theory of musical multimedia with its focus on the way music interacts with visual components.

In addition, the understanding of the music-movement relationship will be discussed on the basis of the psychological theories of perception and cognition so my thesis also has a strong link to the field of music cognition/music perception. It is tempting to add one further research question, which might be formulated as follows:

- *How does our material collected from the sound-tracings and free dance movements reflect underlying perceptual and cognitive processes?*

This would be a question typical of research into music psychology in the way it addresses how musical experience may be understood and interpreted in light of psychological, in this case perceptual/cognitive, theory (Deutsch, 1999).

And we may go a step further and re-phrase the question so that it emerges as a purer psychological research question:

- *Is it possible to explain our collected material with reference to perceptual-cognitive theory?*
- Or alternatively*: to what extent does the proposed perceptual-cognitive theory account for the phenomena observed?*

The aim of this chapter is to discuss how the present thesis is contextualised within the fields of music theory and music perception. This will lead to a clarification of the aims and the methodology of the thesis, as well as its relevance within musicological research.

The first section will present some of the most central themes from musical multimedia studies, whereas the second section will review theory and research that discuss sound-movement relations within the context of music psychology. The perspectives differ with regards to aims and methods, but also with respect to the kind of scientific discourse to which they belong. The review of these perspectives suggests a continuum from an approach to analysis that is basically hermeneutical on the one hand, to psychological research into perception and cognition that employs experimental methods on the other.

The expression *music psychology* refers to the broad research field that examines how psychological processes are involved in the experiencing and performance of music. In my thesis I am particularly interested in the perceptual processes that are involved in the way we *perceive correspondences*. In accordance with *ecological theory* I will understand *perception* as being intimately linked to movements of the body, so that the perception-movement connection constitutes a central element of cognition (Gibson, 1986). This understanding of the relation between perception and movements of the body is of great value within a thesis on music-movement relations. Another core assumption of ecological theory is that perception is closely connected to and enabled by an emergence of meaning, i.e. to perceive is to understand (Noë, 2004). This means that perceptual and cognitive processes are viewed as inseparable. Building on this perspective, my thesis on music-movement correspondences is a thesis on both *music perception* and *cognition*. However, since I am building the thesis on research and theory that are specifically concerned with the role of perceptual processes, I would prefer to refer to the thesis as a study of *music perception*.

2.2. Interpretation of meaning in musical multimedia

Musical multimedia may be broadly defined as an art form in which music is combined with one or a number of other art forms, such as dance, poetry, film images, still images, theatre etc. It may be argued that these quite different forms interact with music in different ways, so that each combination would require a theory of its own, a theory that addresses the specific genre in question. With this in mind, my overview of key issues in musical multimedia does not pretend to be exhaustive. Nevertheless, for illustrative purposes I think it is justified to treat musical multimedia as one genre, and would suggest that there are a few recurrent themes that are relevant across the multimedia genres.

The point of departure is topics presented in Nicholas Cook's *Analysing Musical Multimedia* (Cook, 1998) and Michel Chion's *Audio-Vision* (Chion, 1994). Cook's contribution offers interpretations of different types of multimedia, ranging from commercials and music videos to animated film and opera. In the introductory chapter Cook discusses the combination of music and images in a commercial for the *Citroën ZX 16v*[9]. He suggests that images of the car and a landscape painter are combined with Mozart's music from the *Figaro Overture* so that the car becomes more than a mere mechanical device to bring cargo and people from one location to another. The fusion of music and images makes us view the car as humanised by linking the inanimate to the highly valued performance of art.

> ...its overall message is that the ZX 16v represents an ideal synthesis of art and technology, and the music plays an essential role in articulating this dialectic (Cook, 1998; p 6).

In addition to this introductory example of music-image analysis, the second part of the book is devoted to more elaborate analyses of multimedia works: Madonna's *Material Girl*, Walt Disney's *Fantasia* (with music from Stravinsky's *Rite of Spring*), and finally, Godard's use of Lully's music in *Armide*, (which is a sequence from the film *Aria*). In general, it may be said that these analyses offer interpretations of emergent meaning based on discussions of the way music relates for example to images, narrative and words.

The notion of meaning being *emergent* is central both to Cook's theoretical considerations as well as to the way he approaches analysis. Referring to the Citroën commercial, he claims that music does not possess meaning. On the contrary, music acquires, or offers, meaning as a result of interaction, i.e. Mozart's music becomes meaningful within the context of the images as well as within the context of our historical and cultural contingencies. Moreover, he states that the relationship is reciprocal; the images do things to the music and the music does things to the images.

> If the music gives meaning to the images, then equally the images give meaning to the music. Another way of putting this, [...] is that meaning is constructed or negotiated within the context of the commercial. In which case, instead of talking of meaning as something the music *has*, we should be talking about it as something that the music *does* (and has done to it) within a given context. (Cook, 1998; p 9)

[9] The commercial was broadcast on television in December 1992

This basic notion of musical meaning within the context of multimedia leads to the claim that any analysis of musical multimedia should pay special attention to the interaction between music and images. Cook suggests that the analysis of music and images is not about studying each of the components isolated one from the other; it is about asking questions about the emergent properties of music and images as a perceived interaction. Accordingly, since musical analysis is the analysis of relations between performed single tones, analysis of musical multimedia is the study of how components, such as music, sounds, voices, images, body movements etc., interact; what they do to each other, and the way meaning emerges from these relations.

> To analyse music is to be committed to the premise that music is in some sense more than just a pile of notes; indeed, it is precisely the difference between a pile of notes and a piece of music that constitutes the topic of analysis. But we can be more specific than that. To analyse music is also to be committed to the idea that we perceive the notes in terms of the relationships between them; we perceive each note as influencing, and being influenced by other notes – or at any rate, if we do not, it is hard to see what we could be analysing. In a nutshell, we analyse the interaction between the elements of the music: that is what analysing music means. And exactly the same applies to multimedia. To analyse something as multimedia is to be committed to the idea that there is some kind of perceptual interaction between its various individual components, such as music, speech, moving images, and so on; for without such interaction there is nothing to analyse. (Cook, 1998; p 24)

Another contribution to the topic has been provided by the French theorist and composer Michel Chion (Chion, 1994). He proposes the same general view as Cook that an essential aspect of music in film and other multimedia is that seeing and hearing simultaneously alters each sensory modality reciprocally, and that a major theme in audio-visual analysis is the perceptual experience that emerges as a result of interaction. He suggests the term *added value* to denote the fused experiencing of music and images; the perceptual value that results from integration, i.e. we see things that were not visible without the music; and vice versa, we hear other aspects of music and sounds than we do when listening just to the music without the accompanying images.

In Chion's perspective on musical multimedia, points of synchronisation play a prominent role in the interplay between sounds and visual components. In the synch-points, relational aspects come to the fore, as they demonstrate that auditory and visual components have the ability, when occurring at the same moment in time and at the same location in space, to merge into a perceptual whole, i.e. to be perceived as belonging to the same event. This makes it possible during the editing process to dub the actions of the images, e.g. 'steps in the dark', with a wide range of sonic qualities, thus producing different audio-visual colours and atmospheres. Chion arrives at the term *synchresis* to describe this audio-visual phenomenon:

> *Synchresis* (a word I have forged by combining *synchronism* and *synthesis*) is the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time. This join results independently of any rational logic. [...] Synchresis is what makes dubbing, postsynchronisation, and sound-effects mixing possible, and enables such a wide array of choice in the processes. (Chion, 1994; p 63)

Chion does not enter into a broad discussion of the mechanisms behind the phenomenon; he merely suggests that audio-visual convergence results as a combination of features of the perceptual system and qualities of the sonic and visual components:

> Certain experimental videos and films demonstrate that synchresis can even work out of thin air – that is, with images and sounds that strictly speaking have nothing to do with each other, forming monstrous yet inevitable and irresistible agglomerations in our perception. [...] Synchresis is Pavlovian. But it is not totally automatic. It is also a function of meaning, and is organized according to gestaltist laws and contextual determinations. (Chion, 1994; p 63)

Cook, however, addresses the question of correspondences and differences on a much larger scale. The focus on relation and interaction between musical and visual components as a prominent quality of musical multimedia implies that analysis needs to address the question of the ways in which relation and interaction between the media emerge. He suggests that some kind of correspondence between music and image is a prerequisite for interaction to occur. As a starting point, Cook explores whether synaesthesia[10] may form a basis for establishing a relation, but argues that this kind of correspondence is too static. Instead, he suggests that similarity is on the one hand based on what he calls quasi-synaesthetic correspondences, and on the other, that a relation is established via the way metaphorical transfer enables similarity. By quasi-synaesthetic correspondences is meant that sound and visual events are perceived as resembling each other in certain qualities, such as brightness and size. For example,

> [...] practically everyone – not just synaesthetes – agrees that the sound of a flute in high register is 'brighter' than that of a tuba; conversely, the tuba has a bigger sound – not necessarily louder, but bigger – than the flute. (Cook, 1998; p 75)

A more indirect kind of correspondence may be found in the process of metaphorical transfer. Referring to Lakoff and Johnson's theories about metaphors (Lakoff & Johnson, 1984), Cook argues that when visual and sonic components are brought together, correspondences occur as an emergent property. The audio-visual correspondence is thus a result of the mind's capacity to transfer meaning from one domain of experience to another. In other words, it may be difficult to predict beforehand which sounds will 'go well' with which images as the correspondence emerges as a result of the interaction itself.

As we can see from this brief review, Cook and Chion both place emphasis on the role of interaction; i.e. they are both concerned with the specific qualities or meanings that emerge from the combination of co-occurring sonic and visual components. This

[10] *Synaesthesia* is usually understood to refer to the reported phenomenon that a few people have established a fixed relationship across two sensory modalities, e.g. that a specific smell consistently evokes a specific visual sensation. For most people the connections between the senses are less fixed, and the general capacity to integrate the sensory information from different senses will be referred to as *cross-modality* or *multisensory perception* (Stein, B. & Meredith, 1993). Since *synaesthesia* is understood as a special case of multisensory linkage, the phenomenon will not be further discussed in this thesis.

emphasis on interaction is by no means a new idea, and it has been combined with the notion that music is fundamentally different from other art forms. Since music has been regarded as "different", music has in discussions of musical multimedia been assigned a special role compared to that of visual elements, for example.

This view is found in music philosophy of the nineteenth century with its emphasis on music as "the language of feeling". This was rooted in two interrelated fundamental aspects of music as phenomenon: first, music is not representational; it is not bound to denotations of objects and events in the outer world; and second, music is characterised by its temporal, flowing character; the way sounds seemingly appear and evaporate[11]. As Richard Wagner was introducing the concept of Gesamtkunstwerk in the mid-nineteenth century, he was affected by this emerging philosophy of Romantic music. His vision, articulated in the essay *Zukunftsmusik* (Wagner, 1979)[12] written around 1860, was to renew the opera genre by combining music, poetry, theatre and scenography, in other words all available components on the opera stage, so that the interaction of these elements would create a unique aesthetic experience. Music was assigned a central force in his operatic works; the music's role was to strengthen and accentuate the poetic quality of language as it appeared in the libretto.

A more recent contribution on the topic is found in the book *Composing for the Films* (Adorno & Eisler, 1994)[13]. The authors, Theodor Adorno and Hanns Eisler, claimed that the main role of music in film should be to constitute an antithetic relation to the images. The images are inherently direct representations of the surfaces of objects and events. Music, on the other hand, has the capability to dive under this surface and draw attention to what is going on beneath it. In this way Adorno/Eisler were emphasising the non-representational quality of music. At the same time, they criticise the common practice of Hollywood film-making, i.e. that composers of film music in this genre merely illustrate what is going on in the images, and thus do not make use of the potential in a synthesis of two opposing media.

As noted above, what the Wagner and Adorno/Eisler accounts have in common is the idea that music has special qualities that distinguish it as an art phenomenon from other arts. The American film theorist Edward Branigan in the essay *Sound, Epistemology, Film* elaborates a similar view (Branigan, 1997). Branigan examines on the basis of theories of perception, or what he refers to as a phenomenology of perception, the way sonic and visual elements are perceived differently. His starting point is to question the view that the relation between sound and images as they are combined in a film montage is basically non-hierarchical: "Film theorists may have moved too quickly in recent years in declaring the sound track of the film an equal partner with the image in some grand democracy" (Branigan, 1997; p 95). He bases his argument on an examination of the phenomenology of sound as opposed to vision, i.e. the way he intuitively perceives differences in sonic and visual events.

---

[11] The emergence of these ideas is discussed in Andrew Bowie's essay *Music and the Rise of Aesthetics* (Bowie, 2002)
[12] The essay was translated into English by Robert Jacobs under the title *Music of the Future*, published in 1979
[13] First published in 1949

> A phenomenology may provide important clues to what we believe sound to be. Sound and light may have the same physical basis in wave motion, but they are perceived differently. Lightness and colour appear to reside *in* an object – to be the quality of the object – rather than emanate *from* an object. By contrast, we think of sound as coming from a source, from an object: a radio, a door, a boot. Colour is (seemingly) possessed, but sound is made. Thus we tend to hear sound as transitory and contingent – an on/off phenomenon – while vision is more absolute (a reference-point, if you like). (Branigan, 1997; p 95)

Branigan suggests that the referential quality, apparently inherent, in our perception of visual objects gives vision precedence over auditory perception. He claims that this hierarchical order is embedded in our culture, but that it may also even be grounded in biology: "The way in which human biology is able to exploit the physical difference between light and sound results in sound having a lesser survival value for humans than light" (Branigan, 1997; p 96). With reference to Christian Metz (Metz, 1985) he suggests that sound relates to the image as an attribute, i.e. as an adjective or verb to a noun:

> This fact [the survival value of vision] may underlie the claim by Christian Metz that sound is basically *adjectival* while vision is a *noun*. Metz addresses the epistemological issue of sound by arguing that all perception derives from the *naming* function of verbal language. When we see a 'lamp' and can name it, the identification is complete and all that could be added would be merely adjectival - a 'tall, reading' lamp. When we hear and name a sound, however, the naming remains incomplete. A 'whistling' sound still needs to be specified: the whistling *of* what? *from* where? The whistling *of* the wind *in* the trees *from* across the river. (Branigan, 1997; p 96)

In my view, if we imagine a film scene, e.g. a car crashing, the visual sense informs us about the *what*-aspect of the event, a 'car crash', whereas the sound colours the event with intensity and emotional content. Thus, the two senses complement each other; they take two different roles. And they may be seen as relating to each other in a hierarchical manner, as long as we accept the idea that naming an object has precedence over valuing it in emotional terms. Moreover, identifying the two senses as fundamentally opposing channels of perception paves the way for understanding the relationship as antithetical, in Adorno/Eisler's terms.

This understanding seems plausible if we look at the practice of combining sound and images in film. For example, we all know that scenes of 'steps in the dark' may be combined with a wide range of sonic, step-like qualities, and that the sound editing process in this way is used to add emotional content to the scene. Moreover, it would appear that sound and music are much more flexible with respect to the ability to move between the different narrative layers of the film. Music is commonly used non-diegetically, i.e. it is not directly related to the primary level of narration. But there are also examples of music alternating between being background music and what is known as "source" music (i.e. the musical sound emanates from a source visible on the screen such as a radio or an orchestra)[14]. Moreover, what is also very common in contemporary film is to merge diegetic sound and non-diegetic music so that the sounds of the actions on the screen become elements in the musical texture. In accordance with Branigan's views, we may suggest that this flexibility results from

[14] One such example of music alternating between narrative layers is found in the opening of Stanley Kubrick's *Eyes Wide Shut* and his use of Shostakovich's *Second Waltz* from *The Second Jazz Suite*.

the basic transitory, non-permanent character of music and sounds, whereas the fact that the images appear to a lesser degree in this fluctuating manner in film practice is probably due to the more permanent quality of visual objects.

2.2.1. Core themes of musical multimedia: the role of interaction *and* the question of similarities/differences between media

To summarise this section, the musicological discourse regarding musical multimedia tends to fluctuate between concrete interpretations of multimedia works *and* discussions of music as a phenomenon, i.e. the way music is perceived in an interplay with other art phenomena. As will be outlined in this section, one main theme is the question of interaction; i.e. that multimedia is characterised by the way the visual and auditory change each other reciprocally, so that synergy effects emerge. Another central theme represented by the contributions provided by Adorno/Eisler and Branigan is the issue of how the relationship between music and images should be understood on a principle level; is the relationship hierarchical, complementary or dialectical? This kind of discussion focuses on the differences between music and other media. In contrast, we have the perspectives offered by Cook and Chion who pay more attention to the role of audio-visual correspondences (e.g. on the basis of metaphor as discussed by Cook) and convergence (e.g. the role of synch-points as emphasised by Chion).

Although visual objects and events are more immediately associated with 'names' and 'labels', it should, in my opinion, also be considered that objects are characterised by their movements in space, and by the way forces are involved in these movements. If we imagine a scene with a car running into a tree, or an actor falling to the ground from a tall building, and if we imagine these events without sounds, we would most likely be able to experience the 'car crash' and the 'falling body' with vivid emotional content, and not merely as referential 'nouns' without descriptive 'adjectives'. Thus, despite the permanent feature of objects, these same objects become visual events, and they acquire transitory and temporal aspects when they are set into motion.

Second, and based on theories of perception that assign motor imagery a central role in both auditory and visual perception, I would promote the perspective that sonic and visual elements, when understood as time-dependent phenomena, are both characterised by their transitory, non-permanent features. This theory, a recurrent theme in this thesis, proposes the understanding that seeing and hearing are perceived in the same 'world' of experience. This is quite contrary to Branigan's way of reasoning which seems to rest on the assumption that images and sounds are perceived in two separate, parallel "worlds" of experience.

And finally, the idea that vision might have a greater survival value than auditory perception in terms of biology is challenged by research into multisensory perception. The central claim of this research field is that evolution has provided our perceptual system with a means of integrating simultaneous sonic and visual events; and moreover, that it is precisely this fundamental multisensory feature of perception that promotes and enhances survival, not the individual senses operating in isolation.

2.3. Music-movement relations within the context of music perception

In this section I shall give examples of how sound-movement relations are treated within the context of music psychology. In the introduction to the volume *The Psychology of Music* Diana Deutsch states that the aim of music psychology is to "interpret musical phenomena in terms of mental function – to characterize the ways in which we perceive, remember, create, and perform music" (Deutsch, 1999, p *xv*). With this in mind, the question is: how does the sound-movement link affect musical experience in terms of cognitive and perceptual processes? This shifts the focus into a new direction; the aim is no longer to analyse instances of musical media in terms of symbolic meanings, but to discuss characteristics of music perception.

2.3.1. Examples of observed sound-movement relations

A study that is much cited in discussions of the music-movement link was conducted by Kronman & Sundberg (Kronman & Sundberg, 1987; Sundberg & Verillo, 1980). The study compares timing properties of retardation in music performance with a model for retardation in locomotion, such as running or walking. In a selection of recorded music performances (in performances that the authors refer to as "motor music", mostly by Bach), the decreasing tempo towards major phrase endings was analysed and measured. By employing regression analysis an 'average' timing profile of the performances of final ritards could be obtained. The model for deceleration of physical motion built on the pre-supposition that such sequences of slowing down to a full stop are typically performed with a constant negative acceleration. The average timing profile of the music performances were similar to the model for deceleration in physical motion. The findings suggest a link between timing in music and timing of body movements. The authors concluded that "the musical ritard can be seen as an allusion to physical motion" (Kronman & Sundberg, 1987; p 68).

The music-movement link has also been addressed with respect to pitch perception. In a series of studies the Japanese psychologist Mariko Mikumo examined strategies for storing and recalling pitch information (Mikumo, 1998). In one of these experiments she studied a motor-encoding strategy and its effectiveness for re-calling and re-producing short melodies. 25 young, quite experienced, pianists participated in the study. They were asked to listen to a short series of single notes and encouraged to do finger-tapping as though they were playing the melody on a keyboard. Trials using this motor-encoding strategy were alternated with trials which did not employ such a strategy. The performance of recalling the melodies was monitored, and trials with and without finger-tapping were compared using statistical analysis. The analysis suggested that the motor-encoding approach enhanced recall. It seemed that the trained pianists were utilising their knowledge of the keyboard and the way they related finger and arm movements to sounds of different pitches. Through years of practice on their instrument a robust relationship between sound and sound-producing actions has been established. When this intimate link is encouraged and activated in the finger-tapping strategy, recall of the melodies is considerably enhanced, as demonstrated by the study.

2.3.2. The notion of motor program

A central issue in cognitive psychology is to work out theories of the mental functions that may account for observations of sound-movement relations, such as in the studies referred to above. The key concept is *schemata*, i.e. cognitive structures in which previous experience is stored in a format that enables retrieval and performance in the present. For example, it is believed that movement patterns, such as walking, cycling, eating and speaking, are stored in a kind of *motor program*. A *motor program* is a cognitive structure that contains typical features of a movement sequence, a kind of extract that on the one hand enables re-production of a movement sequence, e.g. 'walking', but that, on the other hand is flexible so that the movement pattern accommodates environmental changes, e.g. the surface on which the person is walking (see Rosenbaum, 1991 for a discussion).

Applied to the results of the Sundberg study of retardation, it may be proposed that the perception, and most likely also the performance, of timing in music is linked to the timing properties of a motor program, in this case the biomechanically constrained deceleration profile of a runner who is slowing down. Similarly, this understanding would imply that the perception of pitch, exemplified in the Mikumo study, is intimately linked to motor programs that the participants have established through years of practice.

The notion of a motor program has been introduced in studies of music perception and performance in various ways. Schaffer defines a motor program as "the mechanism that enables the coordination of a movement sequence" (Shaffer, 1981; p 327), and discusses the concept in connection to studies of skilled musical performance. As another example of the way the concept has been applied to music research, Clarke incorporates the concept in a generative model of musical expression (Clarke, 1993). In a study of the way pianists move while playing a Chopin prelude, it is suggested that movements that are not directly related to sound-production, such as body swaying or head nodding, may serve the purpose of structuring the performance temporally, and that these movements may reflect an underlying motor program (Clarke & Davidson, 1998).

2.3.3. The role of motor imagery in music perception

Godøy has suggested that music perception may be conditioned by a link between the way motor programs are involved in music production on the one hand, and the imagery of motor actions (sound-producing movements) while listening to music on the other (Godøy, 2001). The key theoretical notion of my own study is that listening to music activates imagery of motor actions, and that this *motor imagery* is an integral part of the perception of sonic events, such as musical processes. As noted in chapter 1 it is believed that this *motor imagery* is rooted in the way the listener imagines the movements, or aspects of movements, that might produce the musical sound. Thus, the movements people do when listening to music, nodding the head, swaying the body or tapping a foot, may be understood as traces of this covert imagery; as movements of the body that capture aspects of the way we imagine the source of the sonic event.

To elaborate on this notion we (the members of the *Musical Gestures Project*) carried out a session of video recordings of air-playing, i.e. moving the arms, fingers and the body as though playing an imagined instrument in the air. This is an activity that children occasionally include in their play, but it is also quite a serious matter for those adults who participate in competitions of air-guitar playing.[15] Within the context of a discussion of musical imagery, the phenomenon concretely illustrates our ability to simulate actions that we imagine are involved in a musical performance. What is also apparent is that we are able to do this without any specific training on an instrument, or to put it another way, when we see a child playing air-guitar, we recognise the allusion to guitar-playing, although the movements may be quite approximate compared to actual guitar-playing.

The study of air-playing attempted to demonstrate that even people with little or no training on a musical instrument can utilise knowledge of sound-movement relations, *ecological knowledge*, when attending to music. For this reason, both professional musicians, musicians on an intermediate level of expertise, as well as novices were included in the study (Godøy et al., 2006b). Air-playing the piano, guitar and drums was video-recorded, and air-piano playing by five respondents was chosen for a more detailed examination. We analysed the audio-visual material with respect to correspondences between the air-playing movements and five different musical examples. The analysis was based on qualitative judgements performed by the research group (see chapter 7 for a further explanation of the analytical procedure). Although the experts demonstrated a greater degree of precision, coarse correspondences could be observed on all levels along the expert-novice continuum (Godøy et al., 2006b). The tentative conclusion to be drawn from this is that the mimicry of sound-producing actions is quite spontaneous and requires little or no special training.

In a discussion of the role of motor imagery, the less precise responses of the novices are interesting. Rather than viewing them as imperfect performances due to little or no training, they may be understood as gestural 'sketches' that serve the purpose of actively exploring and 'making sense' of the piano excerpts. Based on the understanding provided by *motor theories of perception* (see chapter 3) these explorative sketches, understood as reflections of covert motor imagery, are assumed to originate from an ecological knowledge of sound-movement relations.

By introducing the *motor-program* concept, music researchers are acknowledging the strong link between mental mechanisms that control and coordinate motor actions and the way we memorise, perform and perceive music. This brief introduction to the concept of motor program suggests that there are variations in the way the concept is applied and understood. For some researchers, *motor program* refers to a rehearsed sequence of music and the way it is stored as a succession of motor commands in the memory of a musician. Used in this manner (e.g. Shaffer, 1981) *motor program* is understood as pre-programmed schema that the musician utilises in new performances of the same piece, relatively unaffected by feedback during the performance.

In contrast, we have the understanding of air-playing exemplified in the air-piano study, reported above. Here the movement sequence emerges on the basis of a pre-

[15] See web: http://www.omvf.net/2004/ilmakitara.php

understanding of sound-movement relations, as well as the feedback that the performer receives from the music. Since the movement is continuously affected and modified during the performance, the expression *motor program* may be misleading, in that the term *program* may be associated with something that is prescribed in advance (see Rosenbaum, 1991; pp 108-110, for a discussion).

2.3.4. Studies of interaction effects and audio-visual match/mismatch

Researchers of music perception have also been concerned with audio-visual *interaction effects*, i.e. the way auditory and visual stimuli may modify each other when combined. For example, two studies, the first carried out by Vines and colleagues (Vines, Krumhansl, Wanderley, & Levitin, 2005), and the other by Krumhansl and Schenk (Krumhansl & Schenk, 1997), have attempted to examine how audio-visual interaction may influence the following elements:

- *grouping*, i.e. how separate tonal elements are tied together so that they are perceived as a temporal unit (e.g. how a stream of successive tones is perceived as a melody line (more about this in chapter 5), and
- *intensity*, i.e. how musical processes may appear with different levels of tension or energy (e.g. the way a musical phrase may be perceived with an increase-decrease pattern in terms of intensity, more about the concept of *intensity* in chapter 4).

The question addressed in these two studies is whether the perception of segmentation and intensity are changed when the musical process is combined with a visually perceived movement, e.g. whether the temporal boundaries of a musical segment are displaced as a result of audio-visual interaction, or whether parts of the audio-visual compound are perceived with more or less intensity compared to when the music or the movement is heard or seen alone.

There are also examples of studies that address the issue of interaction effects, but where the stimulus material has been simplified (e.g. the visual component is geometrical figures) (Marshall & Cohen, 1988; Sirius & Clarke, 1994). Here the dependent variable is defined as the observer's attitudes towards geometrical figures, i.e. how the attitude towards a visual component is changed when the auditory component is changed.

Another study of this kind includes the question of *match/mismatch* between auditory and visual components. One example is a study performed by Iwamiya, in which the respondents were asked to judge match/mismatch in 40 different audio-visual instances (Iwamiya, 1994). The first 20 were taken from commercially produced material, some of them consisted of filmed performances of musicians, and others were combinations of music and scenes, such as landscapes, computer graphics and natural phenomena. They were called matched stimuli, according to the judgement of the producers. The next 20 examples were labelled mismatched stimuli; they were either obtained by combining videos with 'wrong' audio tracks, or by delaying the soundtrack of videos of performing musicians.

Audio-visual match/mismatch has also been studied experimentally with respect to geometrical shapes and synthesised sounds (Lipscomb & Kim, 2004), dance and music (Mitchell & Gallaher, 2001), and musical excerpts and the imagined movements of a human character (Eitan & Granot, 2006). These studies may be seen as closely related to studies of *multisensory perception*. Thematically, research into *multisensory perception* is concerned with how the sensory modalities, such as vision and audition, are integrated so that they work together to make sense of objects and events. One effect of this integration is that it seems that the senses in many cases modify each other; the McGurk-effect and the ventriloquist effect are perceptual effects (to be further explained in chapter 3) that demonstrate multisensory interaction (McGurk & MacDonald, 1976; Stein & Meredith, 1993). Methodologically, research into multisensory perception is affiliated with a research tradition that applies quantitative, reductive approaches. In light of this, music research that addresses interaction effects and audio-visual match-mismatch is connected to such studies of multisensory perception both thematically as well as methodologically.

2.4. Discussion

There are considerable differences with respect to both aims and methodology between the perspectives that have been presented in this chapter. In interpretations of art works, as exemplified in the first section, a hermeneutical approach is applied, characterised by a fluctuation between discussions of subjective experience and the outline of different types of pre-understandings, such as historical, aesthetical, psychological and socio-cultural. In the studies of music perception, on the other hand, the methods may be seen as being inspired by a methodology that originates from fields of research outside the humanities, e.g. psychology. Despite the differences, there are two themes concerning audio-visual relations that seem to overlap across the perspectives; the first concerns *interaction*, the second relates to questions of *similarities and differences*.

2.4.1. Interaction effects/emergent meaning

Regarding the first theme, *interaction*, both music theorists and researchers in the field of multisensory perception take a special interest in the way our perception of sonic events are changed when sound is interacting with sensations to other sensory modalities. The music theorist Cook speaks of *emergent meaning*, suggesting that combining music and sounds with other components of multimedia, e.g. images, results in qualities that are specific to the audio-visual composite, and which were not there prior to interaction. Similarly, the study of interaction effects is one of the main themes of multisensory research.

The question of interaction effects is an important theme in my own thesis since correspondence may be understood as an emergent feature resulting from multisensory interaction. This will be further discussed on a theoretical basis in chapter 3, and this understanding will be taken into consideration in the analyses performed in the second part of the thesis.

2.4.2. Similarities and differences

Second, we may identify a common theme in the question of differences between the sensory modalities. For example, Branigan postulates a fundamental difference between audition and vision, and proposes that the resulting epistemological differences constitute meaning in musical multimedia. Cook, on the other hand, works out a model in which both similarities and differences are incorporated, thus he is suggesting that meaning emerges as an interplay between what is perceived as similar and what is perceived as different in music and images. For this model of musical multimedia to work, it is required that the media find some kind of meeting point. Both the motor theory of perception as well as studies of multisensory perception provide psychological understandings of, and explanations for, the way the senses and the accompanying actions are intimately linked. Moreover, motor theory proposes that motor imagery is involved in both auditory and visual perception. So, on the one hand we have theories that argue that the senses are intimately linked through similar perceptual processes, and that they are linked because of their main and mutual aim to provide functional perceptions of external objects and events.

In contrast, we have theories that claim as a starting point that audition and vision exist in different worlds of experience. In connection to this, it should be noted that although studies of multisensory experience primarily focus on integration, they do take also into account that there are certain differences between the senses, e.g. that audition is characterised by a higher temporal resolution, whereas vision has a higher spatial resolution. In multisensory research this understanding is referred to as the *modality-appropriateness hypothesis* (Shams, Kamitani, & Shimojo, 2004) (see chapter 3). The ventriloquism effect is explained by this hypothesis. When the ventriloquist minimises his or her own lip movements while slightly exaggerating the lip movements of the puppet, it appears to the audience that the sound is coming out of the puppet's mouth. The effect occurs when visual stimuli manipulate and override the way we locate sound in space, i.e. since vision seems to dominate audition in spatial tasks, the perceived location of the sound is altered from the ventriloquist's mouth to that of the puppet.

As pointed out in chapter 1, the main objective of the present thesis is to identify and describe correspondences between simultaneous gestural and musical processes as such correspondences appear in the empirical material from the perspective of an observer/perceiver. Thus, the thesis is thematically contextualised within the presented perspectives on musical multimedia through the focus on *similarities* and *differences*. The theoretical point of departure is that music and movement are perceived in the same 'world' of experience. This would suggest that there is common ground provided by cognitive/perceptual processes (to be further discussed in chapter 4); that sound and movement relate to each other as a perceptual reality (Clarke, 2005). This means that, despite the differences between audition and vision, I shall build on the notion that the perception of correspondence constitutes a reference point on the basis of which musical multimedia are interpreted.

To illustrate this point, I would like to return to the example that was introduced in chapter 1. Merce Cunningham's composer Earle Brown describes a dance performance in which a single high-pitched tone on the violin co-occurs with a big leap performed by Cunningham. In the example Brown mentions, there are two contrasting kinds of relations between music and movement; first, the mickey-moused relation often used in cartoons, and second, the instances of "happy accidents" that often occurred in Cunningham's performances; the first one planned and based on music-movement mimicry, and the second one non-planned and emerging from music-movement divergences. It is a basic notion of this thesis that the experience of both Disney's "mickey-mousing" and Cunningham's "happy accidents" may be grounded in an expectation, or a pre-understanding, of the ways music and movement belong together. In mickey-mousing this expectation is fulfilled. In the example sketched out by Earle Brown, we can imagine that the accompanying violin tone fully or partly violated the expectation.

It seems to me that a notion of similarity is introduced in Brown's own interpretation of the "happy accident". Saying that the violin tone differs from what might be expected, e.g. a heavy chord (or a "great, big 'oooooah' music" as suggested by Brown), implies that he has a conception of what might sound similar to the great, forceful leap. From this it is tempting to suggest that a pre-understanding of similarity is involved although there is no correspondence; that the audio-visual relation is evaluated on the basis of a pre-understanding of similarity. Furthermore, we may, rather than seeing music-movement relationships as based either on similarities or differences, approach the audio-visual compound as a mixture of both similarities and differences. For example, we may speculate that the high-pitched violin tone vaguely resembles the great leap kinematically, whereas they differ in dynamics, i.e. the forces required to perform the leap are heavier and more abruptly applied, contrasted to the continuously applied and weaker force involved in producing the violin tone. The analyses of audio-visual examples in the second part of the thesis will be performed accordingly, since I shall attempt to describe both similarities and differences in music and movement.

2.4.3. Interest in emergent meaning vs underlying mental functioning

The core themes concerning audio-visual relations are treated differently by the hermeneutical approach compared to multisensory research. The hermeneutical approach is concerned with interaction effects as a phenomenon that makes an important difference in the way we experience musical multimedia; interaction effects fascinate by virtue of their unpredictability and the diversity of meanings they potentially bring out. In multisensory research, interaction effects are viewed as a fundamental feature of cognitive functioning and perceptual processing. Thus, analysis of musical multimedia may on the one hand concentrate on interaction effects as meaning, or expression, as such, or on the other hand on the underlying processes that constitute these meanings.

This would suggest that theory development plays a central role in multisensory research, as it seems that one of the main objectives is to observe and measure interaction effects systematically, and on the basis of experiments formulate theories of multisensory perception. This means that audio-visual phenomena are not interesting by virtue of their expressivity and how they are interpreted. Their role is to verify or support hypotheses about cognitive/multisensory functioning, and observations and measurements are valued in terms of how they are able to explain phenomena. In light of this, research within the field of music psychology, e.g. studies of the music-movement link, seems to have much in common with multisensory research, as in many cases it focuses on theories of mental functioning (cf. the quotation about music psychology from Deutsch, 1999; p *xv*). For example, multisensory research may conduct experiments on the ventriloquist effect in order to provide evidence for the modality-appropriateness hypothesis**.** Similarly, the air-piano thesis mentioned earlier may be viewed as an attempt to support the idea that motor imagery is an integral part of auditory perception.

This thesis is written so that the first part reviews theory, whereas the second part concerns analyses and discussions of the collected empirical material. The question is: how do theory and analyses relate to each other; what is the role of theory in this thesis?

Doubtless, with cognitive/perceptual theories that in different ways discuss the sound-movement link, the thesis is contextualised within music cognition/perception. However, it is important to make it clear that the main objective of the thesis is not to provide evidence or support for psychological theories about mental functioning. The thesis will concentrate on describing music-movement correspondences as they appear as observed phenomena, rather than explaining correspondences in terms of cognitive processing; and the thesis may be characterised as qualitative and explorative as it attempts to capture and describe a wide range of aspects within which music and movement may correspond. he following summarises the role of theory in the present thesis:

- The theory serves as a basic understanding of music-movement relations: the view that sound and movement are intimately linked through perceptual processes provides a theoretical basis on which the idea of correspondences as reference points in musical multimedia is grounded.
- The theory offers an understanding of the *flexibility* and *non-arbitrariness* of music-movement correspondences.
- The role and characteristics of non-symbolic aspects of meaning are elaborated on theoretically. This also entails a discussion of embodiments of meaning and the interplay between the symbolic and the non-symbolic.
- Importantly, theory provides a basis for discussing different approaches to the analysis of music-movement relationships and leads to clarifications of analytical concepts which are implemented in the analytical procedure
- Finally, the way the sound-movement link is theoretically approached in this thesis allows sound-movement relations to be viewed as a kind of pre-understanding, in a hermeneutical sense, which constrains musical experience

2.4.4. Zoom level/reduction

If we compare Cook's analyses of interaction effects with the methods applied in multisensory research we can also see a marked difference with respect to the level of complexity in the observed phenomena. Cook, on the one hand, takes into consideration a full musical texture, including melodic, harmonic, rhythmical and timbral aspects, and considers the way this compound of elements evolves in time;

and moreover, that interacts with images of equal complexity. In his analysis of the Citroën commercial, the meaning is interpreted on a symbolic level with connotations of "art" and "technology" as socio-cultural references; and it is suggested that meaning emerges as a result of a dialogue, or a negotiation, between music and images, and as a result of many contributing layers.

Studies of multisensory perception, on the other hand, often zoom in at quite a low level of meaning and examine integration effects within the limited time-window of a few seconds, and stimuli used are sometimes very simple, e.g. using sinusoidal tones and geometrical two-dimensional figures as independent variables and using a judgement of direction of apparent motion as a dependent variable (see chapter 3).

A study conducted by Lipscomb and Kim (Lipscomb & Kim, 2004) exemplifies how such a reduced approach may be applied in an investigation of audio-visual correspondences (to be further explained in chapter 3). Single tones from a synthesizer played in succession are used as auditory stimuli, which are combined with geometrical figures as visual stimuli. Furthermore, the stimuli are characterised by the way they change one feature at a time. The sonic examples are characterised either by pitch change, timbral change or by a change in loudness. Likewise, the visual examples differ from each other with respect to changes in shape, size or colour. These are combined in different audio-visual combinations and respondents are asked to judge the extent of match versus mismatch.

The advantage of this kind of approach is that it allows for a systematic investigation of whether, for example, feature A in music is perceived as 'more' or 'less' corresponding to visual feature B compared to visual feature C. This is in accordance with the basic idea of experimental methodology: to reduce or condense a phenomenon so that a limited number of variables may be systematically varied, and so that the observation and measuring of effects on a dependent variable are facilitated.

In the Lipscomb and Kim study, none of the sound examples have changes in two or more features at the same time. However, this is not unproblematic since musical sound is characterised by the way many different features co-evolve in time and reciprocally influence each other. Based on the understanding of music as a multidimensional phenomenon, I shall later elaborate on the assumption that simultaneous changes in, for example, three features such as pitch, loudness and timbre may afford different kinds of movement responses; in other words, it is suggested that the flexibility of music-movement correspondences may be discussed in light of this multidimensionality. Considering that one of the main objectives of my own thesis is to describe both the flexibility and non-arbitrariness of music-movement correspondences, I would prefer to use sound examples that involve co-evolving features.

In the observational studies that provide the empirical basis for this thesis, we have chosen relatively short sound examples with different levels of dimensionality. In the sound-tracing study we used excerpts that were quite simple, e.g. a short trumpet sound with no change in pitch, loudness or timbre, and examples that were characterised by simultaneous changes in more than one feature, e.g. pitch and timbre. For the video-recordings of free dance-movements we used excerpts that were even more complex in terms of co-evolving features.

So, on the one hand we have the requirement for a certain degree of multi-dimensionality to allow for an analysis of flexibility. On the other hand, the examples have been carefully picked out and edited so that each example is characterised by changes in a limited number of features. This has been performed to allow for systematic observation, but reduction has not been performed to the degree illustrated by the Lipscomb and Kim study.

To sum up this discussion, the present thesis on music-movement correspondences is contextualised within both music theory/analysis and music perception:

- The thesis is connected to a kind of music theory/analysis that investigates the way musical elements interact to produce emergent features, and may be seen as an extension of this approach as the analyses are concentrated on the way musical elements work together with gestural elements.
- In this sense the thesis is similar to Cook's approach to musical analysis (cf. Cook, 1998); however, the thesis differs in its limited focus on non-symbolic aspects, whereas Cook includes symbolic aspects in his analyses.
- The emphasis on non-symbolic aspects, such as dynamics/kinematics and chunking, pulls the thesis in the direction of music perception.
- Finally, the thesis is connected to music perception in the choice of its theoretical framework, as well as in the way methodological elements from experimental research are employed.

2.4.5. Relevance and applications

To conclude this chapter I would suggest that this thesis is relevant on a theoretical level to the following:

- The thesis is affiliated to the *Musical Gestures Project* with its "primary objective […] to work towards *a coherent theory of the relationship between musical sound, human movements and musical concepts* […]"[16]. Thus, the present thesis may be seen as a contribution to working out a general theory of sound-movement relationships.
- Since *ecological theory* and *motor theories of perception* are applied as a theoretical foundation, the thesis may be seen as a contribution to the music-perception field as a whole, as well as a more direct follow-up for example to Eric Clarke's application of ecological theory to musical analysis and theory in his book *Ways of Listening* (Clarke, 2005). Clarke has proposed that the music-movement relationship may be viewed as a perceptual reality and is not solely characterised by metaphorical transfer. This notion is central to this thesis.
- Finally, since the thesis focuses on audio-visual correspondences, it is taking part in the general discussion of musical multimedia theory, e.g. the role of similarity between music and other forms of articulation

The thesis may contribute to future studies of music-movement relations with respect to the way music-movement instances are analysed:

[16] http://musicalgestures.uio.no

- first, in the elaborations of **analytical concepts and procedures**,
- second, in the **annotated audio-visual data**, i.e. in the concrete and detailed descriptions/analyses of sound-tracings and video-recorded dance movements,
- and third, in the methodological issues that are raised.

To my knowledge, there are only a limited number of studies that have examined music and human-body movement in the way I do in the chapters 6, 7 and 8. True enough, the Vines et al and the Kruhansl and Schenk studies referred to earlier in this chapter, as well as a study of the way the visual perception of performing musicians may affect the musical experience (Davidson, 1993), are examples of research that use music and body movement as empirical material; however, they do not address the issue of correspondences specifically. The Lipscomb and Kim study (Lipscomb & Kim, 2004) does systematically discuss match-mismatch between sound and visual impressions, but the stimulus material (simple synthesised sounds and simple geometrical, animated shapes) differs markedly from the kind of empirical material on which my own discussions build.

There are two studies that seem to approach correspondences in music and body movement in the way the present thesis does. The first one is a match-mismatch study of music and dance with a large number of participants (N=942); university students were asked to detect matches between a piece of music and dance movements, but this study does not discuss on which basis in terms of musical and gestural features the judgment is made (Mitchell & Gallaher, 2001). The other study, on the other hand, is much more specific with regards to the way music and movements may match, e.g. with respect to changes in tempo and dynamics, and this study attempts to observe correspondences systematically (Eitan & Granot, 2006).

In conclusion, the present thesis builds on research of the kind exemplified by the last two papers mentioned above. It should be added that an important feature of my own thesis is that I do not solely assess correspondence in terms of an overall match/mismatch in each audio-visual example; the aim is to describe fluctuations in the degree and quality of correspondence within a time-window, and to do this with reference to musical and gestural features. This focus on the audio-visual process is also to some extent found in the Eitan & Granot study (cf. audio-visual mappings with respect to changes in tempo and dynamics).

My objective is to contribute to a more detailed knowledge of the way correspondences are perceived in music-movement relations. The annotated audio-visual data and the analytical concepts and procedures that are worked out in later chapters may be applied by fields of research and practice in which music and movement are a major ingredient. One example is the field of music technology that explore how to build new interfaces that map movements and sound[17]; another example is the practice of music therapy in which the sensitivity to music and movement plays a prominent role for the interaction between therapist and client (Pavlicevic, 1990; Trondalen & Skårderud, 2007).

[17] See e.g. http://www.cost287.org/ and http://www.nime.org/

Chapter 3. An understanding of sound-movement relationships based on theories of perception and cognition.

> We are told that vision depends on the eye, which is connected to the brain. I shall suggest that natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system (Gibson, 1986; p 1).

3.1. Introduction

The main purpose of this chapter is to work out an understanding of sound-movement relations based on theories of perception and cognition. There are two premises underlying this: first, the assumption that listening to sound and seeing a movement are intimately linked by fundamental features of perceptual and cognitive processes. Second, as a perspective on musical multimedia, the understanding that will be worked out here is grounded in the idea that these perceptual processes do not only apply to sound and movement in everyday experience but also to music-movement relations in artistic expressions. Both everyday sounds and musical sounds are produced by movements, and it is assumed that any kind of sound evokes motor imagery. This implies that similar perceptual processes are involved when attending to everyday sounds and musical sounds. This also suggests that knowledge of the way sound and movement 'usually' appear together constitutes a pre-understanding that contextualises the way we experience music-movement relations.

In light of the empirical material used in this thesis, i.e. the sound-tracing task during which the participants were asked to make a drawing movement on the basis of a single sound, and the study in which the dancers were asked to move freely to a musical excerpt, the following questions are brought to mind:

- How does sound lead to a motional response? This question might be asked in different terms, e.g. how does sound *specify*, *guide*, *suggest*, or *prescribe* movement; or to rephrase the question in light of ecological theory: how does sound *afford* movement?
- Which aspects of sound do the participants in the sound-tracing study and the dance-movement study attend to; how do these aspects of sound make them perform movements in a certain manner?
- How may these immediate responses to sonic events, as demonstrated by the empirical material, be understood in light of theories of perception and cognition?
- Or stated otherwise: when we see someone kick a ball, we know that what we see and what we hear is one coherent event, and we 'see-hear' the event as one immediate moment without needing to think about it; similarly, when we see a dancer performing a rhythmic movement, we may in some cases perceive the accompanying music as having a similar rhythmic flow, as though the music and the dance movements were perceived as one coherent, fused audio-visual event.
- These examples beg the general question: how can these observations, the immediacy and intimacy of the sound-movement link, and the sensitivity with

which we attend to such phenomena, be understood in light of theories of perception and cognition?

- Finally we may return to and rephrase the opening question about how sound leads to movement: how might other sensory processes, such as vision, proprioception, haptics etc., affect the perception of sound?

The first field to be explored is research into multisensory perception. The main idea of this field of research is that perception is fundamentally multisensory, which means that the perceptual system, which consists of different sensory modalities, such as vision, audition, proprioception, olfaction and smell, is designed so that the integration of the senses is facilitated. This means that the sensory modalities work in concert to make sense of events in the environment (Calvert, Spence, & Stein, 2004). Accordingly, it is believed for example that the synthesis of hearing the sound of the ball being kicked and seeing the kick occurs before the event is broken up into two single events, one sonic and one visual. It will be argued that a study of perception should include this perspective, suggesting that studying one sensory modality at a time would omit prominent aspects of perceptual processes.

The second part of the chapter will address the role of multi-sensory integration in a wider sense. Over the past decades, *sensorimotor* processes have by some theorists been assigned a central role in perceptual experience. It is proposed that the way sensory and motor processes are interwoven and reciprocally affect each other is fundamentally constitutive for perception. The integrated sensorimotor pattern of seeing and sensing one's own body movement makes up what we often refer to as a visual experience. In this chapter the role of sensorimotor processes will be elaborated on with reference to three inter-related theories: the *ecological theory of perception*, the *enactive approach*, and *motor theories of perception*.

These theories will lead to a discussion of auditory perception, i.e. what does it mean for the experience of sonic events that perception is constrained by ecological knowledge about the sound-movement link; and how can the initial question about how sound *suggests* or *affords* movement be understood within this theoretical framework?

Based on the theories reviewed, I shall outline an understanding of the way perception is linked to action. This will form the basis on which the analytical approaches in chapter 6, 7 and 8 are grounded. The main philosophical view proposed is that perception is an active and explorative process that draws on a "reservoir of sensorimotor skill" (Noë, 2004; p 27). In light of this, the performance of movements as observed in the studies of sound-tracings and the free dance-movements may be understood as the explorative behaviour of perceivers attending to a sonic event. In other words, the dancers derive meaning from the sound by making a movement. The explorative movement is based on some kind of motor imagery, which again is rooted in a sensorimotor skill.

In the theories on cognition presented in this chapter, sensorimotor experience is considered to constitute a pre-requisite for the emergence of meaning. Therefore, as an epilogue to the chapter I shall briefly review the view proposed by Mark Johnson who suggests that sensorimotor experiences and the way they are stored in the memory is an important source for our use of metaphors (Johnson, 1987). His theory of metaphor, and thus also of abstract thinking, bridges the gap between non-conceptual experience and conceptual meaning.

3.2. Multisensory perception: how the sensory modalities affect, complement and modify each other

Many of us have enjoyed the experiences offered in a 180º cinema. Although standing upright on the floor, we get the feeling of actually being on a roller-coaster ride; what we see on the wide screen gives us the same feeling of being bumped to the side in a curve, or the sensation evoked when the roller coaster falls steeply. The visual sensations give rise to bodily sensations because the screen is surrounding us so we cannot receive correcting visual stimuli.

Experiences of this kind are the starting point for research that examines the way the senses are intimately linked. Based on the theoretical notion that perception is fundamentally multisensory, the multisensory field, by applying methods from experimental psychological research, has aimed to analyse in detail the way the senses are integrated. The multisensory feature of the perceptual system implies that the senses are working together to derive meaning from events; they complement each other; they change and modify each other; and this interaction affects the way we perceive the foreground and background on the perceptual scene.

Another essential aspect of this multisensory property of perception is that sensory processes are integrated with and constrained by the actions we perform during the perceptual act. This means that hearing and seeing, as well as smelling and touching, are integrated with our sensations of the way the body moves, the kinaesthetic sense. This means that to fully understand perceptual processes, aspects of multisensory integration have to be accounted for. Researchers of multisensory perception justify this view by pointing out the following:

- First, so far no organisms have been discovered in which the neuronal system is organised so that the senses are completely separate from each other (Stein & Meredith, 1993).
- Second, studies of neuronal activity in animals have revealed that areas in the brain have been specially assigned to integrate multisensory stimuli (Stein & Meredith, 1993). The neuronal basis for multisensory integration has received much attention in recent years (Calvert, 2001; Calvert, Brammer, Bullmore, Campbell, Iversen, & David, 1999; Calvert et al., 2004).
- Third, research into cross-modal processes in infants suggests that the ability to integrate appears very early in life, which has led to the assumption that cross-modal integration is a feature innate to our perceptual system. For example, two/three-week old infants can imitate the facial expressions of the carer, meaning that the visual sensation of ‘what they see’, e.g. the adult opening his or her mouth, is transferred into motor activity as well as the proprioceptive monitoring of ’what they do’, the baby opening its mouth (Stein & Meredith, 1993). Furthermore, infants are puzzled when simultaneous visual and auditory events are presented non-ecologically; e.g. they are confused when the location in space of the mother’s voice is manipulated experimentally, so that the visual perception

of the lips moving does not emanate from the same position as the sound of the voice (Bower, 1974). Young children also demonstrate amodal perception in that they are capable of "translating" features, e.g. the intensity of a sound, into similar features of a movement or other kinds of visual events. This capacity is considered to be of crucial importance for the development of early communication (Stern, 2000). In more recent contributions, amodal perception, often referred to as multisensory redundancy, is considered to be a pre-requisite for the development and maturation of the perceptual system, seen both from a behavioural as well as a neuronal perspective (Lewkowicz & Kraebel, 2004; Wallace, 2004).

"Everyday" experiences such as those demonstrated in the180º cinema draw one's attention to multisensory processes. We may think of a number of other experiences of this kind. For example, when we see someone kicking a ball far away, we become puzzled for a moment that the sound of the foot kicking the ball is delayed compared to seeing the kick, as we are used to 'seeing-hearing' such an event as a simultaneous audio-visual event. Yet another effect of the senses working in concert occurs when we try to fathom what a conversational partner is saying in a crowded room with other noises and other people talking. Under conditions like this, a common strategy is to attend simultaneously to both the speech sounds and lip movements, so that lip-reading strengthens the auditory input. This phenomenon, often referred to as the "cocktail-party effect", the ability to attend to one single voice in a noisy room, has been demonstrated experimentally by comparing the number of words perceived with and without using the lip-reading strategy (Calvert, Brammer, & Iversen, 1998). The effect has been explained by findings from studies of neuronal activity in animals. It has been shown that the activation of multimodal neurons (i.e. neurons that respond for example both to auditory and visual inputs) is considerably increased by simultaneous audio-visual stimuli compared to when the auditory and visual components are presented alone. In an experiment, the response of what was termed a multimodal neuron in a feline brain, i.e. a neuron that is capable of responding to stimuli from more than one sensory modality, was measured. Figure 3 shows how the response was considerably weaker when the cat was presented visual and auditory stimuli separately compared to a combination of the same visual and auditory stimuli (Stein & Meredith, 1993). In fact, neuronal activation seems to be multiplied by integration, referred to as a supra-additive effect, and does not emerge as a mere summation of the effects of auditory and visual stimuli (Calvert et al., 1998).

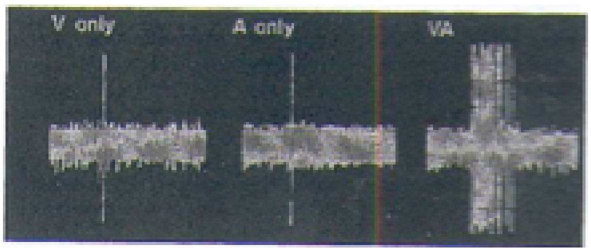


Figure 3. The figure illustrates how the strength of neuronal response is enhanced when the auditory and visual stimuli are presented in the audio-visual condition (Stein & Meredith, 1993).

Other behavioural effects that have been discussed include research into the ventriloquist effect. McGurk and his colleagues demonstrated that speech sounds can be altered by visual lip movements, so that we hear a different sound when combined with visuals compared to when the sound is heard in isolation (McGurk & MacDonald, 1976)[18]. The ventriloquist effect occurs when visual stimuli manipulate and override the way we locate sound in space (Stein & Meredith, 1993; p 1). When a ventriloquist minimises his or her own lip movements while slightly exaggerating the lip movements of the puppet, it appears to the audience that the sound is coming out of the puppet's mouth. The effect emerges as a result of an interplay between the following conditions:

- The synchronisation of lip movements and sound, i.e. whether the lip movements appear to resemble features of the voice
- The distance between the ventriloquist and his puppet, i.e. if the distance is too large the effect ceases to be effective and the location of the sound source moves back to its actual position

3.2.1. Modulations of vision on audition and vice versa

From this brief introduction to the field I shall now proceed to review research that addresses details of audio-visual interactions, and then discuss to what extent the findings may contribute to our understanding of music-movement relations. Although both the neuronal processes that underlie cross-modal integration as well as issues concerning the development of multisensory perception in early life are important themes of my research field, these topics will not be addressed any further in the present thesis. Instead I shall concentrate on the behavioural effects of cross-modal processing, focusing mainly on audio-visual interaction.

Audition may modulate vision in various ways (Shams et al., 2004). Regarding temporal perception in vision, experiments would seem to point in the direction that simultaneous sounds may both affect the perceived duration of a visual stimulus (Welch & Duttonhurt, 1986) and temporal resolution. For example, Fendrich and Corballis (Fendrich & Corballis, 2001) demonstrated that a visually perceived flash is perceived earlier when it is preceded by a sound, and perceived later when it is followed by a sound. Concerning the inter-related aspects of stimulus intensity and attention, studies indicate that visually perceived intensity is strengthened by a simultaneous sonic signal (Stein, London, Wilkinson, & Price, 1996) and that introducing a sound can support the identification of a target visual stimulus that is presented in a series of distracters (Vroomen & de Gelder, 2000). Also with respect to intensity perception, Kitagawa and Ichihara (Kitagawa & Ichihara, 2002) illustrated that a seemingly looming visual shape affected the perceived dynamics of a sonic event. As visual stimulus material, they used the shape of a square that was animated so that it gradually increased in size, thus giving the impression of visual motion in the depth plane. This stimulus was presented to participants so that they adapted to the visual stimulus. When the display was combined with a steady, non-changing tone, it appeared to the observers that the tone was increasing slightly in intensity.

[18] For a demonstration, see web: www.media.uio.no/personer/arntm/McGurk_english.html

Cross-modal effects are reported to alter event properties and to produce illusory effects. In one experiment the participants were presented an animated display on which two identical discs were moving towards each other from either side; they then crossed in the middle of the display, each disc proceeding to the opposite side. The majority of the observers reported that they could see two discs streaming through each other. However, if a brief and sharply articulated sound was introduced just before the two discs crossed, the discs appeared to bounce off one another, moving back on the trajectory to the position from which they had started (Sekuler, Sekuler, & Lau, 1997).

In another study an illusory flash was produced by beeps (Shams et al., 2004). One single flash was presented accompanied by two brief sounds, the first beep sounding just before the visual stimulus and the second just after. Depending on the temporal conditions, e.g. the beeps and flash were required to occur within a time-window of approximately 100ms, the observers reported that they could perceive two flashes instead of one.

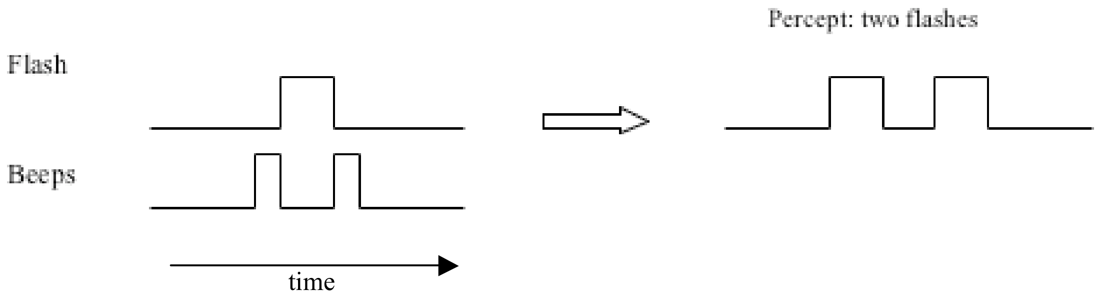


Figure 4. One flash combined with two beeps results in the visual perception of two flashes (Shams et al., 2004)

Another study demonstrated a similar illusory effect: two flashes, both accompanied by a beep, were presented on different spatial locations so that they produced the effect of apparent motion. If a third beep was introduced after the first flash and before the second, i.e. in the middle of the temporal segment, participants thought they could see a third flash lighting up between the two real flashes (Kamitani & Shimojo, 2001).

It has also been attempted to investigate the effects of multisensory integration involving one or both kinds of stimuli including motional characteristics, such as perceived trajectory and velocity. The purpose of this was to examine how time-dependent phenomena may be affected by multisensory integration. The experiment with the bouncing discs that was just described above is one such example.

In another experiment, addressing the question of influences on perceived trajectory, Mateef and his colleagues (Mateef, Hohnsbein, & Noack, 1985) demonstrated that when observers were presented a stationary central sound and a visual shape moving across a screen, the sound appeared to be moving along the same path as the visual stimulus. This seems to be another variant of ventriloquism, i.e. the spatial location of the visual event dominates and overrides the spatial location of the auditory event. A further example of this effect is found in films when the steps of an actor are seen moving across the screen "followed" by the sounds of steps, although the sound has not been panned accordingly in the sound-mix.

Also related to the perception of trajectories, potential interactions were studied in an experiment that included two apparent motion streams, one sonic and one visual (Soto-Faraco & Kingstone, 2004). Auditory and visual apparent motion was presented under four different conditions: same direction-synchronous, conflicting direction-synchronous, same direction-asynchronous, conflicting direction-asynchronous. Participants were asked to judge the direction of the auditory stream. Their judgements were correct under all conditions, except when the condition was conflicting direction-synchronous; this suggests that vision captured the sound motion, or alternatively, that the visual stimuli interfered so that the participants made random guesses. Under the asynchronous condition, there was probably no integration, i.e. the auditory and the visual streams were perceived as separate events.

Integration effects have also been reported with respect to velocity perception. Manabe and Riquimaroux (Manabe & Riquimaroux, 2000) investigated the influence of sound on perceived velocity in a visual apparent motion task. Using a display of three LED-flashes (see figure 5), they asked subjects to compare the velocity of the motion from the left LED to the centre LED (standard velocity) with the velocity from the centre LED to the right LED (comparison velocity). The timing of the flashes was kept constant. Sound bursts of differing loudnesses and durations were presented either just after the flash from the left LED or just after the flash from the centre LED. The researchers concluded that the velocity of visual apparent motion was significantly modified by the sound stimuli, and furthermore, it appeared that short sound durations induced the perception of faster velocity.

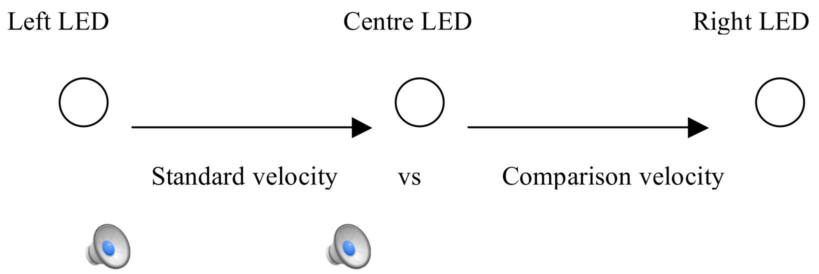


Figure 5. Sound bursts presented just after left LED or just before centre LED made observers report differences in perceived velocity, i.e. between standard velocity and comparison velocity, even though the timing of visual stimuli had not been altered (Manabe & Riquimaroux, 2000).

In sum, these studies suggest that audio-visual integration may influence a number of perceptual features, such as timing, intensity, velocity, direction/trajectory and even spatiality. These are all important features for both sonic and visual events, i.e. events that are temporally, dynamically and kinematically determined. There seems to be a pattern in the way vision modulates audition, i.e. vision typically dominates audition in spatial tasks. For example, with the ventriloquism situation, the perceived location of the sound is spatially moved from the mouth of the ventriloquist to the mouth of the puppet. Conversely, audition seems to dominate vision in temporal tasks, e.g. the timing of a visual flash is changed by the timing of a beep (Fendrich & Corballis, 2001), and likewise, the perceived velocity of visual apparent motion is altered by the temporal properties of an interacting sonic event (Manabe & Riquimaroux, 2000). It is assumed that these observations may be explained by differences in visual and auditory perception, i.e. that vision has a better spatial resolution than audition and that audition has a better temporal resolution than vision. This leads to a more general hypothesis about multisensory integration, the modality-appropriateness hypothesis, which "postulates that the modality that is most appropriate or reliable with respect to a given task is the modality that dominates the perception in the context of that task" (Shams, et al, 2004; p 28).

3.2.2. Findings of multisensory research applied to music-movement relations

On the basis of research into multisensory perception, the following assumptions may be made about how music and movement affect each other when combined:

- We may assume that the musical flow has the potential to change temporal features of the dance movements to be examined. For example, we may expect that the timing of accentuations in the music will override the timing of "nearby" accentuations in the movement.
- Regarding interaction effects on perceived intensity, a further general finding of multisensory research suggests that two relatively weak stimuli, e.g. a weak auditory one combined with a weak visual stimulus, will produce a comparatively strong cross-modal response (Calvert et al., 1998; Stein et al., 1996). The "cocktail-party effect" demonstrates this multisensory effect, in which two temporal patterns, i.e. lip movements and speech sounds reciprocally strengthen each other, thus enhancing perception. It seems fair to assume that sound and movement in a music-movement relationship may also under certain conditions affect each other in terms of perceived intensity; for example, that a relatively weak emphasis in sound is considerably intensified by a synchronised accentuation in movement.
- These two general effects, timing and intensity, may individually or combined have impacts on the way the audio-visual compound is perceived in terms of segmentation, phrasing or overall rhythmical quality. Understood in this manner, the effects of interaction imply a quite direct relevance for the way my recorded material with music and dance combined is perceived as a cross-modal phenomenon.
- Another central theme of multisensory research that seems to be relevant for music-movement correspondences is the question of the conditions that facilitate integration. It seems that integration, and the effects of integration, depend on the interplay of temporal and spatial factors; for example, auditory and visual stimuli that occur at approximately the same time and approximately at the same location in space tend to be perceived as one integrated event, as a perceptual whole, as though the sensations were emanating from the same source. Furthermore, such integrated audio-visual events become foregrounded in perceptual experience, as though the two sensory modalities were strengthening each other. It is also proposed, that in addition to the role of temporal and spatial factors, the nature of

stimuli contributes to integration. This means that the qualities of the sound should approximate, or bear some resemblance to, the qualities of what we see. For example, in the case of the ventriloquist, the emergence of the effect depends to some extent on whether the dynamics of the sound are similar to the dynamics of the lip and head movements of the puppet. Moreover, it is pointed out that the role of similarity does not only apply to isolated moments of synchronisation but also to the way the patterning or ordering of dynamics is distributed within a more extended time-window (Calvert et al., 1998). I will later in this thesis build on the notion that both music and movement are characterised by the way different features, such as dynamics, change over time. Applied to music-movement correspondences, I would propose that when features in music and movement change similarly within the same time-window this will enable perception of correspondence.

So far I have reviewed experiments that are characterised by the way they use simplified auditory and visual stimuli, so that they appear to be laboratory experiments. And I have suggested that the findings from these experiments may be applied to an understanding of music-movement correspondences. To generalise from highly reductive multisensory research to instances of complex music and movement is of course questionable. In many cases the laboratory experiment uses sparse stimuli, such as sinusoidal tones and simple geometric shapes. Furthermore, the audio-visual material is not allowed to evolve over a longer time span, which means that a rhythmical, temporal pattern is not articulated. It should be noted that in most of the experiments reviewed, motion is understood as movement from one spatial location to another, so that the kind of motion that we associate with music, which is more related to dynamical aspects and the temporal patterning of a sonic event, is not addressed. In light of this, generalisations should be made with care; however, I believe that multisensory research is relevant for a study of music-movement relations both in the way the field raises general issues about multisensory perception and in the way it presents specific interaction effects which music research may explore.

3.2.3. Examples of music research into interaction effects

It should be mentioned that two studies within the field of music research, both utilising more vivid, time-dependent, audio-visual material, have attempted to shed light on the effects of interaction. In both studies the participants were assigned three conditions: music alone, movement alone, and music-movement combined. They were asked to indicate perceived segmentation, i.e. the start and end of a phrase/segment/group, and a continuously perceived change of tension or intensity. In the first study conducted by Krumhansl and Schenk (Krumhansl & Schenk, 1997) the audio-visual material consisted of a video-recording of dance choreographed by Balanchine to Mozart's minuet from *Divertimento no. 15*, and in the other study the participants were presented video-recordings of clarinet players performing Stravinsky's *Second Piece for Solo Clarinet* (Vines et al., 2005). In both cases it has been concluded that the temporal patterning seems to be perceived in a similar manner in both the auditory and the visual channel, i.e. the segmentation of the music and the movement is not perceived differently across the three conditions. This means that there are no observed effects of interaction concerning segmentation in these studies. Regarding the question of tension/intensity, the two studies differ in their conclusions. Krumhansl and Schenk find no effect of interaction, whereas Vines and colleagues suggest that visual information, the player's movements, enhanced the perceived tension at some moments and weakened tension at others.

These are two sources of experiments available for discussing interaction effects in which human-body movement is included. In other studies body movement has been replaced by geometric figures as visual stimuli, and interaction effects have been studied with respect to ratings of attitudes, i.e. the studies address another aspect of meaning. Marshall and Cohen (Marshall & Cohen, 1988) presented a group of respondents moving geometric figures and accompanying music with the aim of observing whether the music modified attitudes towards the visual figures. They used the semantic differential judgement technique for ratings of attitudes, according to the dimensions evaluation ('good/bad'), potency ('strong/weak') and activity ('high/low'). Mean ratings for music alone, images alone, and music/images combined were compared. The results suggested for example that strong music (potency) tended to increase the ratings of activity ('higher') and potency ('stronger') under combined image/music conditions. Sirius and Clarke used a similar approach, and they also reported that music had an effect on ratings of computer-generated animated figures (Sirius & Clarke, 1994).

In the described experiments the modulations of audition on vision and vice versa are commonly referred to as *interaction effects*. However, in some cases, researchers have concluded that the observed effects suggest that the audio-visual relationship should be understood as additive rather than interactive (Krumhansl & Schenk, 1997; Sirius & Clarke, 1994). For example, by additive Sirius and Clarke's understanding is that the effect of combining music with images raises the ratings of a specific attribute (semantic meaning) compared to when the images are viewed without music, but the combination does not produce qualitatively different meanings. Accordingly, an interactive relationship is understood as a compound that results in novel meanings, i.e. meanings that were not there prior to combination. This distinction may be useful from an analytical point of view, as it makes clear that audio-visual combinations do not always produce something entirely new. However, I would prefer to use the expression *interaction effect* for any change that is produced by an audio-visual combination. The reason for this is that I assume that modulations ranging from minor adjustments in temporality to considerable changes in intensity will in most cases make a true difference, so that they are not merely changes in degree, but are experienced as changes in quality.

3.2.4. Inter-related aspects of multisensory perception: integration, interaction effects, and event-detection

Multisensory research deals with two parallel themes:

- the question of interaction effects, i.e. in what ways do the senses modify each other when combined?

- the question of conditions that facilitate integration, i.e. how do the temporal, spatial, and patterning features of the sonic and the visual event respectively interact so that the two streams are 'seen-heard' as one fused event?

In my view it is essential to understand interaction effects and integration-facilitating conditions as interdependent aspects of multisensory processes. For illustration we may reconsider the experiment in which a square was animated so that it grew in size and is then combined with a steady sustained tone. The reader may recall that the participants reported that they could hear a tone that increased in brightness. This increased brightness may on the one hand be described as an interaction effect, an example of how vision modulates audition. On the other hand, considering that the interaction effect makes the sonic event appear more similar to the visual event in terms of dynamical patterning, it may at the same time be understood as a condition that facilitates integration.

However, the two inter-related aspects might be viewed from a slightly different angle. Instead of discussing isolated interaction effects, it is tempting to apply a more overarching understanding of cross-modal integration. In the second section of this chapter, from an ecological perspective on perception I shall elaborate the view that perception is basically a process of event-detection. Based on this, multisensory perception may be understood as a feature of the perceptual system that facilitates event-detection; i.e. integration clarifies the phenomenon as an event. So, instead of concentrating on specific modifying aspects the focus is shifted towards looking at the way sounds and movement together constitute an event. Accordingly, the 'bouncing discs' (Sekuler et al., 1997) would no longer be viewed as a kind of mysterious audio-visual illusion, but rather as an example of the way auditory and visual events are combined so that the possibility of perceiving the event in a different manner is being offered, i.e. as a different kind of event ('streaming' vs 'bouncing').

Similarly, in the experiment in which participants were asked to judge the direction of auditory apparent motion (Soto-Faraco & Kingstone, 2004) it seems reasonable to interpret the responses in light of audio-visual, event-constitutive properties. As described above the judgments were correct under conditions in which the combination of auditory and visual apparent motion did not conflict with "normal" audio-visual event properties. The two streams presented in synchronisation and in the same direction apparently caused no problem, and we may assume that the respondents were basing their judgements on a pre-understanding of the way audio-visual events evolve in time, i.e. "when something sonic is synchronised with something visual, they most likely emanate from the same source, and as one fused event they normally travel along the same path in the same direction." Conversely, under the conditions with asynchronous auditory and visual streams, neither was there any conflict because the insufficient synchronicity did not offer the possibility of perceiving the two streams as one single event. This caused the respondents to judge the direction of auditory motion correctly. Finally, applying this kind of understanding, there was a true conflict in event properties when the two streams were presented simultaneously, but in opposite directions, i.e. the event did not evolve "as it should" and the respondents were confused.

What I would suggest here is that the respondents apply a pre-understanding of event-properties when attending to audio-visual phenomena, and that this pre-understanding of the way events to unfold in time constrains multisensory integration. Again the example of the growing visual shape and the increasing intensity of the accompanying tone illustrates the point: It is reasonable to believe that the participants in the study were utilising the pre-understanding based on the everyday experience that a sound emanating from an approaching object, e.g. a car, tends to become louder and more intense.

3.2.5. Perceptual interpretation/event-detection vs amodal perception

The kind of pre-understanding discussed above will later in this chapter be referred to as *ecological knowledge*, and is assumed to play a central role in the perspective that views perception as event-detection. Within the context of multisensory processes, the perspective proposes that integration and interaction effects are constrained by the overall intentionality of perception. This intentionality may be described as the tendency to attempt to put things together so that they appear to be coherent and ecologically plausible events. The case with the tone that increases in brightness due to cross-modal integration may be understood as a situation in which the perceivers add the dynamical feature to the sound to make it match their pre-understanding of temporal events. This suggests that the perceptual process of multisensory integration involves an element of interpretation.

There seems to be a kind of multisensory connection that appears to be more hard-wired, i.e. a mode of matching that is relatively direct and automatic. In multisensory research this capacity of the perceptual system is referred to as *amodal perception*. The expression refers to the observation that something we hear is heard as similar to something we see with our eyes, or do with our bodies. We perceive "something" similar regardless of sensory modality, and this "something" is specific attributes of the percept. The *amodal* attributes are defined as:

> […] those that can be specified equally well across the different modalities, provide the same information regardless of sensory modality, and thus provide the basis for intersensory equivalence (Lewkowicz & Kraebel, 2004; p 655).

Intensity, rhythm, shape, texture, spatial extent, spatial location, duration, temporal rate and rhythm are regarded as attributes that are readily transferred from one modality to another. The ability to perceive amodally seems to be a feature innate to the perceptual system, or at the least a capacity that is introduced at an early developmental stage, and which, as previously noted, is thought to play an important role in the development of multisensory perceptual skills (Lewkowicz & Kraebel, 2004; Lickliter & Bahrick, 2004).

From the viewpoint of studying specific, corresponding features in music and movement amodal perception is important to consider as a source of audio-visual similarities. Amodal perception suggests that we have some kind of shared and relatively predictable idea of the way the intensity of a visual display corresponds for example to the brightness of a timbre, or that the temporal rate of a movement corresponds to the density of onsets in music. Thus amodal perception seems to provide a source for our sensitivity to sound-movement relations that is more automatic, or more determined, than is suggested if we view audio-visual perception as a process of interpretation/event-detection.

An experiment performed by Lipscomb and Kim (Lipscomb & Kim, 2004) may illustrate this duality. They aimed to examine match/mismatch judgements in an experimental design in which single tones were paired with simple geometrical figures. The auditory stimuli varied in terms of pitch, timbre, loudness and duration. Variations in the visual component were based on colour, size, shape and location (on the computer screen). Combining these qualities systematically created a total of 48 audio-visual composites. Participants (N=28) were asked to evaluate the correspondence between the auditory and visual component in each of these samples. The researchers predicted that pitch would be matched with vertical placement, amplitude with size, and that waveform (timbre) would correspond to colour and pattern (shape).

The results confirmed these predictions; the highest ratings were obtained for pitch/location (i.e. high pitch/high placement on screen), loudness/size (big sound/big object) and timbre/shape (smooth sound/round shape, jagged sound/jagged shape). If we look at some of the audio-visual composites that were used, this is not unexpected. As an example of a matching of auditory pitch and visual location, the visual component consisted of a circle changing position from a low, to a middle, to a high position on the screen, and then back again. This was synchronised with the ascending tonal succession of *tonic – fifth - tonic (octave above),* and then descending to *fifth – tonic* (see figure 6).

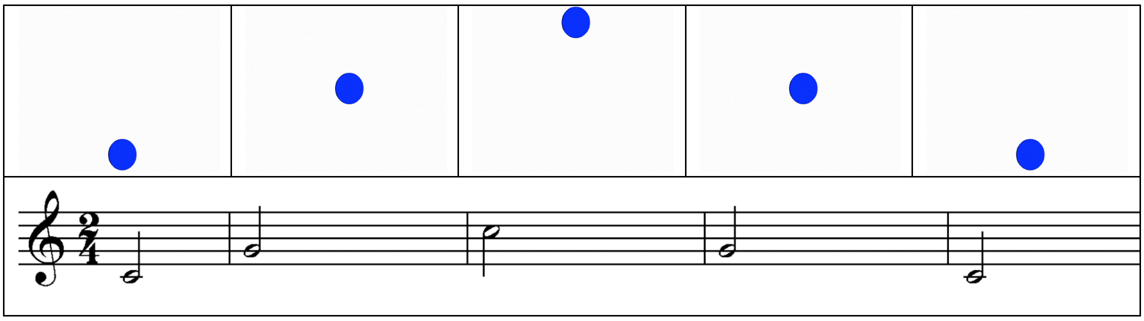


Figure 6. The succession of tones are synchronised with a succession of circles that 'move' upwards and downwards in space (Lipscomb & Kim, 2004).

Apparently, the temporal rate and rhythm are the same for both vision and audition; thus the pairing may be understood in terms of amodal transference. In addition, the change low/middle/high/middle/low in both modalities appears to be another salient feature for a high rating of correspondence; however, the pitch of a sound is commonly considered to be a modally specific attribute and not an amodal feature (Lewkowicz & Kraebel, 2004; p 660). Therefore, it seems that this feature of correspondence must be assigned to another kind of perceptual process.

In another example a succession of tones with the same pitch, played in a *p-mf-f-mf-p* dynamical pattern, is synchronised with circles that increase and decrease in size (see figure 7).

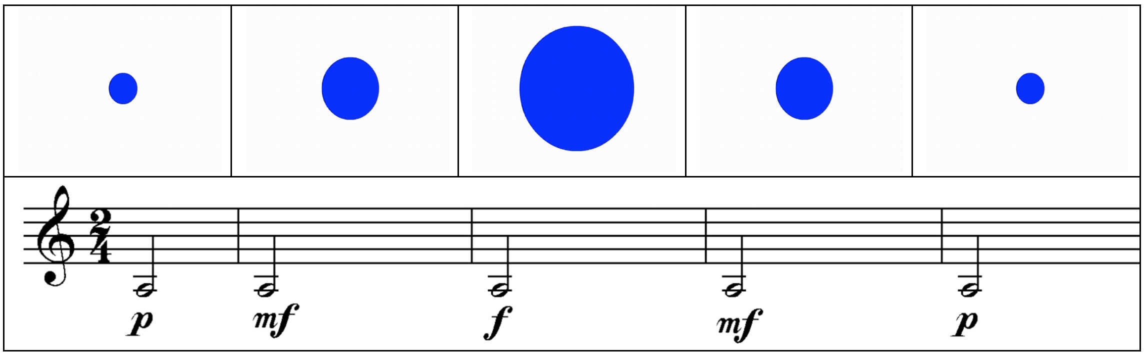


Figure 7. The succession of tones is synchronised with circles that increase and decrease in size (Lipscomb & Kim, 2004).

Again, the temporal rate is similar in both streams. Additionally there seems to be an amodal connection in the change of intensity, i.e. the increasing/decreasing loudness of the auditory component results in an increasing/decreasing intensity profile that corresponds to the increasing/decreasing intensity provided by the small/big/small size of the visual figures. This last kind of correspondence may however be interpreted slightly different. If we think about how sounds are produced, a big body movement, i.e. a movement involving the big parts of the body and relatively large extensions of limbs, normally results in a relatively loud sound. Considering this, the correspondence between a loud sound and large figure may be related to the event-detecting feature of perception rather than to amodal transference. This kind of discussion will be broached later in the thesis.

3.2.6. Preliminary summary: specific effects and automatic vs interpreted correspondence/integration

I believe that two themes are particularly important for theoretical as well as analytical discussions of music-movement correspondences: first, the reviewed laboratory experiments point to the specific effects we may expect from audio-visual interaction. One example is that it seems that the temporality of sounds may override the temporality of the visual stream. Another effect concerns perceived intensity as it seems that auditory and visual streams have the capacity to strengthen each other, e.g. at certain accentuated moments.

Second, the integration of two sensory streams may be approached from an angle which involves the connection being understood as automatic, thus resulting from the relatively distinct features of the two streams. The phenomenon of amodal perception points in this direction, i.e. it has been observed that people seem to have a straightforward idea of how a sonic event, for example in terms of its intensity, temporal rate and rhythmical flow, may be translated into some kind of visual correlate. On the other hand, by interpreting interaction effects as an event-detecting feature of perception, we apply a more overarching "rule" for correspondences. What we may understand as the intentionality of perception would lead us to a consideration similar to the following one: "anything auditory and anything visual that occur at the same time (approximately) and at the same place (approximately) will be explored for similarities to see if there is any possibility of experiencing them as one fused audio-visual event". In this case audio-visual correspondence is understood as an emergent feature of integration.

3.3. Ecological theory: the unity of perception and action and the unity of perceiver and surroundings

As suggested by the term *ecology*, the *ecological theory on perception* views the relation between the organism/perceiver and the environment/surroundings as fundamental to the way we perceive and understand objects and events. It is claimed that perceptual experience and meaning is relative to the way the perceiver has established his/her relation to the surroundings (Varela, Thompson, & Rosch, 1993). This means that perceptual meaning is not given by any physical characteristics of objects and events. Perceptual meaning emerges in the interplay between the perceiver and the surroundings and is constrained by previous encounters with similar objects and events. Understanding perception as event-detection, as was the case in the previous section, is therefore in accordance with an *ecological theory* of perception, since perception is viewed as an explorative rather than as an automatic process.

Another fundamental unity is constituted by perception and action, i.e. action understood as the body movements of the perceiver. The ecological approach assigns perception and action a central role in cognition; as a fused sensorimotor process perception and action are considered the "driving force" of thinking.

The basic notions of an ecological approach have predecessors. One example consist of the theories of perception worked out by the Gestalt psychologists in the early twentieth century, as mentioned in the introductory chapter of the thesis. Furthermore, the view of perception and action as inseparable processes was central in the work of the French philosopher, Maurice Merleau-Ponty (Merleau-Ponty, 1983). In the quotation below he describes the interplay of sensory processes and movements of the body:

> When the eye and the ear follow an animal in flight, it is impossible to say "which started first" in the exchange of stimuli and responses. Since all the movements of the organism are always conditioned by external influences, one can, if one wishes, readily treat behavior as an effect of the milieu. But in the same way, since all the stimulations, which the organism receives have in turn been possible by its preceding movements which have culminated in exposing the receptor organ to external influences, one could also say that behavior is the first cause of all the stimulations (Merleau-Ponty, 1983; p13).

Observations of this kind lead to the understanding of perception and action as depending on each other. Action is guided by perception, whereas perception depends on and is shaped by the actions we do with our bodies as we intentionally attend to events. Such sensorimotor experience is viewed as processes that both structure and nurture cognition.

3.3.1. Sensorimotor processes: perceiving, acting and knowing

By sensorimotor experience we here mean perceptual sensations and simultaneous actions interwoven with the temporal course of each act of experience, as observed by Merleau-Ponty. When bouncing a ball on the ground, we are engaged in a full, sensorimotor experience, including hearing the sound of the ball as it hits the ground, the tactile sensation of the ball touching the palm of the hand, the proprioceptive sensations of our own body moving, and the way we see the actions of our body transmitting energy into the ball: and vice versa, the way we sense the weight and forces of the ball spreading from our hand into the arm and the rest of our body. The sensorimotor experience of playing with a ball is basically multimodal, i.e. a number of sensory modalities are integrated, and these are again integrated with body actions (Stein & Meredith, 1993). Bouncing a ball differs from one instance to the next. The ball may be hard or soft, dry or wet, heavy or light, and the surface we are playing on may vary similarly. We all know that playing with a ball involves the mastering of a skill; what we have learnt is the relationship, the contingency, between the movements of our own body, our sense of attending to the ball with vision, audition and touch, and the material of the ball and the conditions of the surface. And we have learnt it so that we are able to adjust our movements according to varying conditions.

This close relationship between the "player" and his or her body and the environment is emphasised in the *ecological approach to perception* that was proposed by James Gibson. According to this view perception is not something that is going on in the eyes or in the ears alone; perceptual experience is constituted in the relation between the acting animal and its environment.

3.3.2. Environment vs physical world

Gibson distinguishes clearly between the *environment* and the *physical world.* The environment is the events and things in the surroundings to which the organism has established a relation; the surroundings of the perceiving and acting animal (Gibson, 1986). The environment is meaningful to the organism whereas the physical world is not. We do not perceive the physical properties of the ball, its weight in kilograms or its diameter in centimetres. The ball is meaningful through the actions that we perform with it, through our play we know the ball's bouncing and rolling features, as well as its resonant features. The term environment is understood to mean the surroundings of the animal, the objects and events that are acted upon by an animal; and vice versa, an animal is understood as something that cannot be thought of without its surroundings.

> The fact is worth remembering because it is often neglected that the words *animal* and *environment* make an inseparable pair. Each term implies the other. No animal could exist without an environment surrounding it. Equally, although not so obvious, an environment implies an animal (or at least an organism) to be surrounded. This means that the surface of the earth, millions of years ago before life developed on it, was not an environment, properly speaking. The earth was

> a physical reality, a part of the universe, and the subject matter of geology. It was a potential environment, prerequisite to the evolution of life on this planet. We might agree to call it a world, but it was not an environment (Gibson, 1986; p 8)

3.3.3. The concept of *affordance*

This means that perception and behaviour are not directed at uncovering physical properties *per se*; perceptual experience is related to the relation between the environment and organism. To clarify this view Gibson introduces one of the key concepts of his theory, the concept of *affordance* (Gibson, 1986). The environment, in which the organism perceives and behaves, *affords* certain ways of perceiving and behaving. The solid ground says 'walk on me'; the water says 'drink me'; and the ball says 'play with me'. This relational meaning, the *affordance* of an object or an event, which we instantly assign to the things around us, is the result of the organism, with its perceptual, cognitive and motoric systems, co-evolving with its environment. The affordance of a thing or an event is what the environment offers an animal to perceive or act upon and is distinguished from the physical properties of objects and events. Affordance is therefore relative to both the animal and the environment.

> If a terrestrial surface is nearly horizontal (instead of slanted), nearly flat (instead of convex or concave), and sufficiently extended (relative to the size of the animal) and if its substance is rigid (relative to the weight of the animal), then the surface *affords support*. It is a surface of support, and we call it a substratum, ground or floor. It is stand-on-able, permitting an upright posture for quadrupeds and bipeds. It is therefore walk-on-able and run-over-able. [...] Note that the four properties listed – horizontal, flat extended and rigid – would be physical properties of a surface if they were measured with the scales and standard units used in physics. As an affordance of support for a species of animal, however, they have to be measured *relative to the animal*. They are unique for that animal. They are not just abstract physical properties (Gibson, 1986; p 127).

This view implies that perceptual meaning is not *in* the environmental event or objects as an intrinsic property, nor is it solely a function of subjective interpretation. Meaning is constrained by the way the relation between the organism and environment was established in the past and the way this relation is re-created in the present.

In light of this relativity, the relation between man and nature may be understood as quite floating, and perceptual experience as unpredictable considering the wide range of contexts and responses by which events are shaped. However, there are stabilising factors in the system. In both an evolutionary as well as an individual perspective the perceptual and cognitive system have evolved, adjusted and fine-tuned themselves in interaction with nature. Objects of the world appear to behave in certain ways, constrained by physical laws, such as gravitation. Likewise, events unfold in a fairly predictable manner. And our bodies have through experience been attuned to the way objects tend to behave in the environment. These properties, that may be regarded as relative stable, are not properties of the event itself, but are defined by relations between elements of the system, between perceptual processes and external events. The way we know about such events, the way we know how a ball bounces and eventually rolls, and the accompanying sounds when it is thrown down a hill, are denoted as *ecological knowledge*. This kind of knowledge, which is basically relational, supplies perceptual experience with stability, direction and a certain degree of predictability.

3.3.4. Perception as hypothesis-testing

The ecological approach differs from more traditional views of perception in the way the role of perception and action is understood as basically functional. In the traditional view, perception has succeeded when an internal representation of the external world, some kind of internal picture that matches the objects perfectly, is created. From an ecological approach, on the contrary, the role of perception is to establish a relation; perceptual processes should guide action by creating images of the outside world. When I see the ball I imagine the weight, and what it is like to pick it up, and when I touch it with my hand, I imagine what it would be like to lift it, and when I lift it up and feel the weight and the surface material of the ball, I imagine its bouncing properties, and what it would be like to let it bounce on the lawn, on which I have walked on with my bare feet. So, the perceptual images that are evoked from seeing, touching and lifting have an approximate and flexible nature and may be understood as hypotheses, which are tested and modified by the succeeding actions. Perception and actions are thus understood as a process of generating hypotheses and testing them continuously (Berthoz, 2000). Rather than understanding perception as a passive process of uncovering the secrets of objects, perception is viewed as an exploration and interpretation of the relation between the surroundings and myself.

3.3.5. Sensorimotor patterns and knowledge

To play with a ball is a learnt skill, and to imagine the bouncing properties of the ball is also a skill. This means that the present must somehow be linked to past experience. Cognitive psychology has presupposed that some kinds of structures must be part of our cognitive set-up, structures that make past experience available to the present. In playing with a ball, the body will have to adjust to the diversity, resulting from variations in the properties of the ball and the surface. Despite these variations, we are apparently able to transfer past experience to actions and perceptions in the present, so that we do not start from scratch in each new perceptual act.

Building on Gibson's ecological approach, Varela and his colleagues (Varela et al., 1993) have emphasised the importance of sensorimotor patterns, such as bouncing a ball, and have claimed that cognitive structures emerge from these recurrent, slightly varied instances of sensorimotor experiences. In their action-based view of cognition, referred to as the *enactive approach*, they suggest that perception and action supply cognition with structure:

> In a nutshell the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided (Varela et al., 1993; p 173).

What distinguishes ecological theory from more traditional accounts is that this perspective claims that past experience is primarily structured and made available to the present on a pre-conceptual level, rather than in some kind of symbolic form. In line with this, Johnson (Johnson, 1987) has proposed the equivalent terms *image-schemata* and *embodied schema* to denote the kind of structures of mind that contain and extract the diversity of sensorimotor experience. In his view, the schemas are pre-conceptual, meaning that they contain qualities related to the temporal order and dynamics, as well as the multimodal character, of an event. They are extracted versions of general features of the sensorimotor experiences. They are relatively stable, but are continuously modified by new experiences.

With this in mind, sensorimotor experience is understood as basic to our knowledge of objects. Our knowledge of an object is related to the actions we have performed with the object types. This action-based *how* knowledge is viewed as fundamental, and more immediately accessible to the mind than symbol-based knowledge (Lakoff, 1987). This view is contrary to the basic assumption of traditional cognitive theory, in which knowledge of objects is related to the physical properties of the object and the way it relates to other objects in conceptual hierarchies.

The action-based knowledge of objects implies that objects can be conceived differently depending on the context and our actions. The object "ball" may very well be perceived as a kind of chair, as long as the size and shape of it appropriates the action of "sitting down and resting the body". This means also that previous experience does not relate to the present in a one-to-one associative manner. Thus, utilising our learning of how to manipulate a ball is not a matter of re-producing previous actions; it is a matter of transforming images of our body/ball relation, so that a new unique relationship is created. I think that this theoretical understanding accounts well for general observation, i.e. in the way we use objects creatively as tools depending on the demands of the particular situation.

This discussion addresses the relationship between experiences in the past and the present. Within the context of a theory on metaphorical transfer, Johnson proposes that image-schemata, as extracts of previous sensorimotor, multimodal experience, provide an important source for the way meaning emerges in the present. On the one hand, they do this by projecting structure on the present; however, they also allow for a certain degree of flexibility:

> To say that image schemata "constrain" our meaning and understanding and that metaphorical systems "constrain" our reasoning is to say that they establish a range of possible patterns of understanding and reasoning. They are like channels in which something can move with a certain limited, relative freedom. Some movements (inferences) are not possible at all. They are ruled out by the image schemata and metaphors. But within these limits, there is a measure of freedom or variability, that is heavily context-dependent (Johnson, 1987; p 137).

In my view this proposal of the way previous experience enables interpretations in the present may be applied to an understanding of the way pre-understandings of sound-movement relations influence music perception. This suggests that knowledge of sound-movement relations offers a certain flexibility in the way we may experience novel music-movement combinations, but that this flexibility is to some degree limited.

To sum up, I have reviewed approaches to the understanding of perceptual experience, which in their labelling points to two main theoretical ideas:

- First, the term *ecological* implies that perception is constrained by the relationship between the perceiver and its environment.
- Second, the term *enactive* suggests that perception is something that is carried out, that perception involves activity and action.

Seen together this means that perception is understood in light of two inseparable unities: the unity of organism-environment and the unity of perception-action. Within this framework, sensorimotor processes are central in the way they are the basis for cognitive structures on which new acts of perception and new actions are grounded. I have used the action of "bouncing a ball" as an example of this kind of sensorimotor pattern.

3.3.6. Perception as sensorimotor skill

Since bouncing a ball is an acquired skill, it is argued that perceptual processes that are involved in the action also must be understood as a skill (Noë, 2004). For example, to judge the bouncing properties of the ball by touching, seeing or hearing, is linked to our knowledge of how these sensations are interwoven with the movements of the body as well as the behaviour of the ball. This temporal pattern of perceptions, the sense of self-movement not forgotten, and the way the ball bounces on different surfaces is what Alva Noë is referring to with the equivalent expressions *sensorimotor contingencies*, *sensorimotor dependencies* or *sensorimotor skill* (Noë, 2004). In the book *Action in Perception*, Noë discusses for example how the shape of an object is perceived according to the enactive view. From one vantage point, the ball appears to the eyes in a certain manner, but if we move slightly to the side, the appearance of the ball also changes slightly, and if we move the eye, the head or the feet once again, the angle is again shifted, as accordingly is visual sensation. The process of self-movement results in sensory variations to the visual system, but still we perceive the ball as being the same, and as having the quality of "roundness". Thus the shape of the ball emerges in our perception as the result of a pattern of simultaneous variation in the way the ball appears to the eyes and the sense of the self-movement of the body. Through repeated encounters with the object, the perceiver "recognises" similarities in the way the sense of self-movement co-varies with changes in visual sensations over time, so that it appears as an invariant pattern of co-variation. This invariant pattern of co-variation constitutes the sensorimotor skill on which the perceptual experience of "roundness" is grounded (Noë, 2004).

Having established this understanding, one might argue that there certainly are situations in which we look at an object without moving. In such cases there is apparently no sense of self-movement that co-varies with the visual percept, and seemingly the theory falls to some degree apart. Against this it should be noted that, first, a crucial feature of seeing is that the eyes are constantly moving without us being aware of this, i.e. the so-called saccadic eye-movements. Second, support for the role of self-movement may be found in *motor theories of perception*, which pre-suppose that the imagery of movements is an integral part of perception. This implies that seeing the ball involves imagining how the visual sensation would change as a function of self-movements, without actually performing these movements.

3.3.7. The role of sensorimotor experience for music perception and imagery

Based on the general theory outlined so far, which has mainly been illustrated with reference to visual perception, we may suggest that sensorimotor experience also plays a constitutive role in music perception. In the following I will elaborate on a view of music perception which claims that diverse experiences of interwoven sound and movements, such as in sound-producing actions, constitute a pre-understanding of sonic events which is fundamental for the way we interpret music. This perspective includes the notion that imagery of movement is assigned a central role in music perception, in accordance with the *motor theories of perception* referred to above.

The reader may recall the introductory epigraph of this chapter, in which Gibson underscores the way perceptual processes are grounded in the body and environment:

> We are told that vision depends on the eye, which is connected to the brain. I shall suggest that natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system (Gibson, 1986; p 1).

This basic idea would apply to other sensory modalities as well, so that we could rephrase the sentence above by exchanging vision with auditory perception: *Natural audition depends on the ears on the head on a body supported by the ground, the brain being only the central organ of a complete auditory system*. Initially, the term "support" may be interpreted metaphorically, i.e. auditory perception depends on the support provided by the body and the surroundings; our perception of sonic events relies on and is grounded in the relation to the body and environment.

Bregman proposes this notion as an essential understanding in his contribution on auditory perception (Bregman, 1990). Bregman argues that the perception of sound depends on ecological knowledge of sound and the way sound is shaped by the movements and constraints of the environment.

> By a perceptual question I mean one that asks how our auditory systems could build a picture of the world around us through their sensitivity to sound, whereas by an ecological one I am referring to one that asks how our environment tends to create and shape the sound around us. The two kinds of questions are related. Only by being aware of how the sound is created and shaped in the world can we know how to use it to derive the properties of the sound-producing events around us (Bregman, 1990; p 1).

As we attend to sonic events we make use of our ecological knowledge of the way sounds and movements are tied together; i.e. the way movements produce sound and co-evolve with sounds over time, as well as the way different materials result in different sonic features. Analogous to the way an object falls to the ground is predictable to vision because of our knowledge of gravity, sounds are shaped by nature. This suggests that the relation of sound, the body and environment provide constraints that support and shape our perception of sound.

Acknowledging the role of ecological knowledge implies that perceiving sound is something that does not occur in the ears and the auditory system alone. Sound perception is grounded in a pre-understanding of how movements shape sonic events, and how the auditory sensations are contingent on the responsiveness of materials with which our bodies are interacting. With reference to the work by Noë, it was previously suggested that "seeing a ball" and its 'roundness' emerges on the basis of a pattern of co-variation between visual sensations and the sense of self-movement. Analogously, the perceptual experience of a sonic event may be understood as constituted by the way the perceiver recognises a pattern of co-variation, i.e. in the way changes in sonic features and the sense of self-movement of the body co-evolve over time. This senorimotor process is further constrained by the materials involved (wood, metal, the palm of the hand, mallets of different shapes and materials, etc.). Thus, to perceive sound is a skill, a sensorimotor skill that is based on knowledge about sound sources and sound-producing events.

Being equipped with this pre-understanding implies that we are attending to sonic events with certain expectations. When we watch somebody that is going to kick a ball, we anticipate the sound quality, and if it is a ball made of leather, we expect a certain sound quality, and if the sound does not match our expectation we may be caught by surprise and perhaps reason that the ball may in fact have appeared to be made of plastic and not properly filled with air. Our ecological knowledge of sounds, objects and movements implies that we are able to imagine the properties of an object by attending to the sonic features alone. Experiments have demonstrated that the ability to specify objects by their sounds is very accurate (Carello, Wagman, & Turvey, 2005).

The findings of the experiment conducted by Mikumo (Mikumo, 1998), which I reviewed in chapter 2, may be interpreted within this framework of understanding. The participants, being trained pianists, were encouraged to use a motor-encoding strategy, i.e. finger-tapping as though they were playing the melodies, to enhance recall of the short melodic lines. The way the pianists seemed to relate finger-tapping to pitch changes may be understood as a result of their specific ecological knowledge, i.e of the relationship between the movements of their body and the keyboard. The study showed that the coupling of finger-tapping and sound considerably enhanced recall of melodies which suggests that ecological knowledge in this case played an important role in the melody-perception task. The air-piano study, also reported in chapter 2, may be understood similarly. However, in this study the participants demonstrated ecological knowledge in different ways; from the very specific and precise knowledge of the keyboard sound and articulatory finger-movements demonstrated by the professional pianist to the coarse correspondences between movements and dynamical and structural features of the music observed in the novice performances of air-piano playing (Godøy et al., 2006b).

Another important point that should be emphasised is that the role of sound sources in auditory perception leads to the notion that perceiving sound is to perceive it as an event. That is, we do not hear the sound of the ball as an isolated acoustic quality; the perceptual experience of the sound is embedded in the sound-producing actions, so that we hear it as "the sound of someone kicking a ball". Similarly, hearing a melodic phrase played on a piano is not a matter of attending to a series of temporally successive pitches, it is fundamentally a perceptual act of hearing the event of "someone playing a melody with a certain phrasing on a piano with fingers and a supporting body".

If we appreciate this kind of understanding, what happens if the sound source is out of sight, e.g. when we listen to a CD? In this case we do not see the actual event, and we do not see the movements and the materials involved. This has led to the assumption that some kind of motor imagery must be involved in auditory perception, i.e. that we imagine the event and the sound source (Godøy, 2001).

The role of motor imagery is central to theories that are referred to as *motor theories of perception*. One variant of this theory was proposed in the early 1950s and later the theory was explained in an article by the linguists Liberman and Mattingly (Liberman & Mattingly, 1985). They suggest that speech perception is based on the imagery of how speech sounds are produced by movements (lip movements, movements of the vocal tract, control of air-flow, and so forth). In a more recent contribution Alain Berthoz claims that simulation of action is an integral part of any process of perception (Berthoz, 2000). This means that seeing a pen on a table includes simulating the actions of what it would be like to take a step forward and pick up the pen, without actually performing the action, for example so that the perception of distance to the pen is connected to this covert, non-performed imagery of what it would be like to take a step forward. This means that the motor imagery serves the function of predicting the consequences of actions, i.e. to generate a hypothesis about the acting body and the environment.

This line of thinking has found new credibility in the discovery of what have been termed *mirror neurons* (Rizzolatti, Gartilucci, Camarda, Gallex, Luppino, Matelli, & Fogassi, 1990). These are neurons that are activated both when an animal is performing an action, and when the animal is observing the same action being performed. The finding supports the idea that action and perception are interwoven processes, since they seem to employ the same neuronal resources. The discovery of such neurons, as well as further evidence showing that perceiving the body movements of other people results in imitative, motoric activity in the brain (Wilson & Knöblich, 2005) point in the direction that a covert imitation of movements seem to be quite an automatic process, as though we were attending to movements of other people by constantly (and covertly) imitating these movements. In light of this, Wilson and Knöblich propose the notion that a covert imitation of movements is an integral part of the visual perception of biological motion, i.e. that when we see someone walking, motor imagery, in the form of covert imitation, of walking is activated (Wilson & Knöblich, 2005). Where Berthoz uses the term *simulation*, Wilson & Knöblich prefer the equivalent term *emulation.*

> […] we propose here that covert imitation functions as part of a *perceptual emulator*, using implicit knowledge of one's own body mechanics as a mental model to track another person's actions in real time (Wilson & Knöblich; p 463)

Applying these ideas to music perception, it may be suggested that the experience of musical sound is grounded in the way we explore images of the movements that might produce the musical sounds we are attending to (Godøy, 2001) The images are not actual, internal "video clips" of musicians producing sounds on their instruments. They are images of movement qualities associated with sound-producing actions, i.e. approximate images of general features of the way the music might have been produced by actions. If we understand perception as a process of generating and testing hypotheses, I would suggest that motor imagery has an explorative character, i.e. as listeners we explore or "try out" in our imagination the movements that might be involved in producing the sounds. In sum, the view proposed here implies that motor imagery is an integral part of music perception.

Exploring musical processes in terms of motor imagery is assumed to be grounded in previous sensorimotor experiences where movements are interacting with sonic events. We have numerous and diverse experiences of how sounds are produced and shaped by continuous movements, i.e. sound-producing actions related to both natural non-musical as well as musical sounds. I would suggest that all encounters with integrated sound-action events are relevant to our pre-understanding of music and movement relations. This means that we in the perceptual act utilise any available pre-understanding of sound-movement interactions, so that experiences of everyday sounds and musical sounds are both relevant when listening to music. The point here is not the origin of our pre-understanding, but the way we are able to transform previous sound-movement relationships to produce meaning. Hence, any available sensorimotor contingency in which sound is integrated may be utilised, i.e. the psychological process is the same although the category of sound may vary (for discussions of this, see Bregman, 1990; p 460, and Clarke, 2005; pp 71-72).

3.3.8. Motor imagery in perception of music and music-movement relations

On the basis of these assumptions, I would suggest that the movements we spontaneously perform while listening to music reflect the kind of covert musical imagery that is a pre-requisite for music perception. These actions, such as tapping a finger, nodding the head, or stomping a foot, are obviously not actual, perfect re-productions of the sound-producing actions, but they are similar to these actions with respect to more general features, such as temporal order, dynamics and intensities.

Similarly, I would suggest that the movements of the dancers in our study may be understood as a way of exploring the musical examples by performing movements. For example, we may imagine that the dancer, after having heard a few tones, initiates a certain movement, characterised by the speed, trajectory and dynamics that he or she assumes might correspond to the music. The movement may be understood as a hypothesis about the music and the kind of movement that might go well with the music, and furthermore, a hypothesis of how the music might evolve in the near future. As the movement evolves along with the music it may be slightly modified, e.g. in terms of speed and dynamics, so that the observed movements may be viewed as a continuous process of proposing, testing and modifying the hypotheses.

These hypotheses address the relation between the movements of the body and the musical process, and the features in music and movement that make a difference in this relationship. I propose that they may be based on how the dancer imagines that the music he or she is attending to might have been produced by actions. As noted previously, this is not to say that actual images of performing musicians pop up in the dancer's head as video-clips. It suggests that the dancer simulates certain aspects of the sound-producing actions that he or she imagines are involved. These aspects may have something to do with the density of onsets, articulation, speed/tempo and dynamics. This implies that the observed dance movements may be understood as reflecting covert imagery of sound-producing actions, i.e. the dancer projects previously acquired ecological knowledge of sonic events.

The hypotheses put forward by the perceiver are not arbitrary; they are grounded in a pre-understanding of sound-movement relationships. On the other hand, the hypotheses have a sketchy, explorative character. They are not meant to be perfect reproductions of the music, for example in terms of the number of onsets within a time-window. Despite this sketchy character, they will in a number of instances be experienced as resembling aspects of the music, leading to the judgement that music and movement are similar. For example, a high density of onsets in music might correspond to a high density of onsets in the movement pattern although the number of onsets does not match exactly. The kind of similarity that we are speaking of is not a "perfect match" as we have in the case with two identical rows of numbers. It is more like a "good-enough match"; a flexible and approximate kind of similarity, i.e. music and movement are sufficiently close to be judged as similar and not different.

The drawing in figure 8 is meant to illustrate the way I understand the interplay between similarities and differences. The four shapes are clearly not identical, but they articulate both similarities and differences with respect to the aspects of smoothness, number (of peaks) and extension/amplitude (vertical displacement):

- The drawings A, B and C are similar in the way they are smoothly curved, and differ from drawing D
- The drawings A and D are similar in extension and differ from drawings B and C
- The drawings A, B and D are similar with respect to the number of peaks, and differ from drawing C

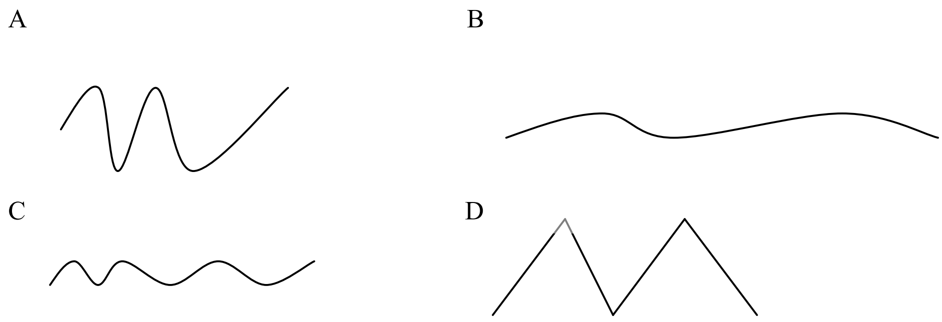


Figure 8. The four different shapes A, B, C and D, are clearly different. As perceivers, we are however able to find similarities. Judging whether they match depends on the features we focus on (number of waves, smoothness of the curves, vertical extension).

Similarly, I would propose that a sonic and a gestural process may at the same time be viewed as both similar and different, i.e. they are similar in some aspects but different in others, and that the perception of music-movement correspondence comprises both similarities and differences.

3.4. Discussion of empirical material in light of theoretical foundations

When collecting the empiricial material for this thesis on music-movement correspondences, we asked the participants both in the sound-tracing and free dance movement studies to make a movement that they thought would best match the sound/music. So the question is: how do the sonic events *specify* or *afford* movement? And, based on the notion that perception is fundamentally linked to a knowledge of production/source, how do the participants approach the task; how may the perceptual process be understood in relation to the motional response?

Considering the tasks in both studies from the perspective of the participants, I would suggest that the movement responses to music may be interpreted within the theoretical framework as follows:

1. *Perception implies a process of approaching "something" in the external world (the environment) with the intention of establishing and exploring the perceiver's relation to the event. This means finding out how this "something" may appear as a meaningful event to the individual perceiver.* This would suggest that the sound-tracer and the dancer perform movements with their body as though asking: "how may this musical event become meaningful to me through my body movements?"

2. *Perception resembles a process of exploration, where the perceiver puts forward hypotheses about possible sound-movement relations and tests them by performing actions.* This means that the participants in our studies, whose intention was to explore the single-tone elements as sonic events, make assumptions about which movements may go together with the musical sound.

These assumptions are reflected in their motional responses, and they may be modified on the basis of what they hear in real-time.

3. *Perception involves anticipating what might be going to happen in the near future, e.g. how the sonic event may evolve in terms of phrasing or dynamics, or how the trajectory of a visually perceived movement may proceed from the 'now'*. Hearing the sound segments as a sonic event, as well as approaching them with a strategy of hypothesis-testing, means that the perceivers in our study initiate and elaborate on their responses by predicting "the next step" of the music. For example, the perceiver may anticipate the dynamic profile of a sonic event, and "think" something like: "it seems that the music is building up tension, this will probably go on for a while, and then we will reach a climax which is followed by a release of tension". Based on this kind of anticipation the perceiver initiates a movement with a similar dynamic profile.

4. *Perception is a sensorimotor skill*, i.e. the dancer or the sound-tracer anticipate how the sonic event might proceed on the basis of knowledge about sound-producing actions, e.g. knowledge about how an increasingly forceful movement tends to bring out an increasingly louder sound. This implies that exploring a sonic event and establishing a relation to this environmental event by making movements with the body draw on previous sensorimotor experiences where sound and action is integrated. In other words, the dancers and the sound-tracers apply an understanding of sound-movement relations when approaching the task. This is in accordance with the view proposed by Alva Noë that "to perceive is not merely to have sensory stimulation. It is to have sensory stimulation one understands" (Noë, 2004; p 181). I would propose that this kind of acquired knowledge of music-movement/sound-movement relations may be viewed as a *pre-understanding* that constrains musical experience. This is a kind of *sensorimotor pre-understanding* that is working in concert with and integrated with other kinds of *pre-understandings,* such as socio-cultural and symbol-based *pre-understandings* of music-movement relations.

5. *The goal of a perceptual process is not to provide a perfect mental representation of the external world, but to create an image of the event that is sufficiently 'good', in the functional sense, to enable and guide action.* To produce a movement with the intention that this movement should correspond to the sonic event does not mean that the movement and the music are required to be identical or the same. It suggests that a wide range of movement variants, i.e. variants that approximately resemble aspects of the musical sound, may establish a relation between the body movements and the music. In light of this, an observed music-movement correspondence may include both similarities and differences, for example, as noted previously, music and movement may on the one hand be coarsely characterised by a high density of onsets (similarity), although they do not exactly match the number of onsets within a time-window (difference).

This list of propositions about perception and music-movement relations serves the purpose of summing up the main ideas presented in this chapter. The main theoretical notion that emerges is that the sound-movement link should be understood as a perceptual reality. In his book *Ways of Listening* Eric Clarke argues that the psychological processes that underlie any kind of sound perception, of everyday as well as musical sounds, make the experience of movement in music appear as a real experience, and not merely as a metaphorical transfer (Clarke, 2005). This view will be elaborated on in the next chapter (chapter 4).

Furthermore, the outlined theory points in the direction of a duality that seems to "stick to" the way we experience music and movement relations. I have formulated an understanding of this duality by considering the way audio-visual integration, or correspondence, may on the one hand be viewed as a relatively direct and robust connection, and on the other hand be considered a relation that emerges on the basis of perceptual exploration and interpretation.

In my view, the *flexible* but *non-arbitrary* nature of audio-visual relations may be understood in this light: the way repeated and varied experiences of how sounds are produced by movements constitute our pre-understanding of the relations between sound and movement. This pre-understanding supplies the audio-visual relation with stability. In addition we have the correspondences that emerge on the basis of *amodal perception*, i.e. the seemingly hard-wired intersensory phenomenon demonstrated when features such as intensity, rhythm, duration and number, are readily transferred from one sensory modality to another. I would suggest that both an experience-based pre-understanding of sound-movement relations as well as an *amodal perception* contribute to the *non-arbitrariness* of music-movement correspondences.

On the other hand, the sound-movement link is perceived flexibly. The notion of *image-schemata*, proposed by Johnson (Johnson, 1987), implies that previous experience is stored to allow for a certain degree of flexibility. The *image-schemata* suggest a range of possibilities, within which new experiences may be interpreted. Our interpretation of a new sound-movement relation may be understood as an elaboration, or transformation, of previous experience, which is done in order to make what we already know fit within the context of the present. In this transformation, which may be seen as a creative process of the mind, lies the flexible and approximate nature of music-movement correspondences.

In this thesis I approach music-movement correspondences from two slightly different perspectives: first, we have the "inside" perspective of the participants in our observational studies as performers of correspondences; and second, we have the "outside" observers of music-movement instances as perceivers of correspondences. I have not attempted to draw a clear line between the "performer-of-correspondence" and the "perceiver-of-correspondence" perspectives. The reason for this is that I understand them as overlapping perspectives. This is grounded in *motor theories of perception* that view perception and production as intimately linked: as an inside-performer I explore the movements that might go together with the music; and as an "outside" perceiver I base my judgement of correspondence on a covert simulation of what it might be like to be a performer of the observed movements in relation to the particular music. In both cases, the judgement of correspondence is connected to how the music-movement combination may be perceived as a coherent audio-visual event.

In chapter 4 I shall elaborate on a theory of the sound-movement link. It has been proposed that the motor imagery that is activated in music perception may be primarily related to dynamical aspects, i.e. the temporal distribution of "forces", of the sonic process. Thus the main theme of the next chapter will be a discussion of *dynamics* and its twin concept of *kinematics* and the different ways these concepts are applied to understanding of music and movement.

3.4.1. Epilogue: the perception-action loop as a the driving force of thinking

In the introduction to the section on ecological theory I stated that one of the main ideas of this theory is that the perception-action loop constitutes the driving force of thinking; recurrent sensorimotor patterns lead to the formation of cognitive structures, which again shape new perceptions and actions. Considering the rather mundane example of sensorimotor patterns that I have referred to throughout the theoretical review, i.e. the action of bouncing a ball, one might ask how this simple action can have anything to do with higher-order thinking. The philosopher Mark Johnson provides a theory that links hands-on experience with abstract thinking.

He refers to the way children typically relate to the world by acting on it; by using their own bodies and senses they explore objects and events while monitoring what is going on in their own bodies. The result is a pool of sensorimotor experiences, i.e. integrated cross-modal units of perceptions and actions which are stored in memory. Importantly, one of Johnson's main claims is that such experiences are stored not as symbolic forms/schemata, but in a pre-conceptual form, i.e. as a kind of mental abstraction for which he proposes the expression *image-schemata*, as discussed previously.

In Johnson's own words the *image-schemata* is "our way of having a world" (Johnson, 1987). This means that children engage in all kinds of actions to explore and 'have a world': they crawl under a table and then out again, repeating this again and again. And they balance on stones or on a log, again repeatedly without getting tired. Relating to the world in this manner seems to be their primary mode of experience. Now, another main point in Johnson's theory is that this mode of experience does not cease to exist, thus criticising Piaget's idea that body-based thinking in early stages of life is replaced by abstract thinking when we reach adolescence. Johnson claims that we bring with us these patterns of sensorimotor experience and transform them so that they can be utilised in abstract thinking. This can be observed in our use of metaphors, e.g. when we are abstractly speaking of the balance in a relationship between two people, this is rooted in direct experiences of balancing our own body (for a discussion of the metaphor *balance* see Johnson, 1987; pp 74 – 98). This means that there seems to be a link between the bodily actions and abstract/metaphorical meaning. Issues concerning the emergence of meaning with a focus on the blurred boundary between the *symbolic* and the *non-symbolic* will be discussed in chapter 5.

Chapter 4. Music and movement correspondences in terms of dynamics and kinematics

> Effort is visible in the action movement of a worker, or a dancer, and it is audible in song or speech. If one hears a laugh, or a cry in despair, one can visualise in imagination the movement accompanying the audible effort. Rudolf Laban in *Mastery of Movement* (Laban, 1971; p 24))

4.1. Introduction

The purpose of the chapter is to discuss the concepts of *dynamics* and *kinematics* so that they can be applied and further elaborated on in an analysis of music-movement correspondences. The reason for discussing these concepts is the hypothesis that music and movement correspond in terms of similar and adequately synchronised changes in music and movement, and that these changes may occur in dynamical and/or kinematical features of music and movement. The experiment described in the previous chapter, which involved a shape increasing in size and the increased brightness of a tone heard, illustrates simply the way correspondence emerge on the basis of similar changes in features; a visual change characterised by its "increased size" is matched to a sonic change characterised by its "increased brightness". The role of *changes in features* is central to the thesis:

- First, time-dependent phenomena such as music and movement are thought to be fundamentally characterised by changes in one or more features. The term *change* is used to mean the way the level/degree of a certain feature may vary within a time-window, e.g. the loudness going from soft to louder, or the density of onsets increasing from the beginning towards the end of an excerpt. The term *features* refer to e.g. dynamical and kinematical aspects of movement or music, as noted above. In this chapter also the features termed *activation* and *effort* will be discussed. Later in the thesis a number of features will be further explained and applied to analysis, such as *density of onsets*, *loudness, pitch range, timbre* and *articulation* for music, as well as *density of onsets, extension, involvement, speed* and *articulation* for movement.
- Second, correspondence is assumed to emerge as a result of the way features in music and movement appear to change similarly within the same time-window.

I shall first review two kinds of theories which both address changes in features. The review will concentrate on the core concepts of these theories: *activation contour* and *effort*. The American psychiatrist Daniel Stern proposed the term "activation contour" as a basic phenomenal feature of non-verbal utterances (Stern, 2000)[19]. An activation contour is introduced in his research as a central component of early, non-

[19]Daniel Stern's book *The Interpersonal World of the Infant* was first published in 1985.

verbalised communication between the caregiver and infant, and is as such embedded in a broader theory of child development and pre-verbal meaning.

Based on studies of the movement patterns of dancers and workers, the choreographer and dancer Rudolf Laban introduced *effort* as a central feature of all human movement, and included this understanding in a comprehensive framework for movement analysis (Laban, 1971; Laban & Lawrence, 1947).

It seems that both Stern and Laban understand their concepts as related to *dynamical* features, i.e. to the forces that cause a movement and the way a movement appears with different levels of intensity (see definition below). Laban does this explicitly when he states that every "human movement is indissolubly linked with an effort, which is, indeed its origin and inner aspect" (Laban, 1971: p 21). Stern's concept of an *activation contour* is connected to dynamics and he discusses how an activation contour is experienced as a temporal pattern in which the pattern is characterised by "dynamic shifts" or changes in intensities over time (Stern, 2000; p 57). To exemplify his point, Stern suggests that this mode of experience may best be "captured by dynamic, kinetic terms, such as 'surging', 'fading away', 'fleeting', 'explosive', 'crescendo', 'decrescendo', 'bursting', 'drawn out', and so on" (Stern, 2000; p 54).

Despite this focus on the dynamical aspects of movement, it is clear that the *dynamics* are closely linked to the *kinematics*. As we can see from the above, Stern proposes what he refers to as "dynamic, kinetic terms" since some of these metaphors, e.g. "fading away", "fleeting", "drawn out", at the same time allude to changes in intensities (the *dynamics*), as well as movement in space (the *kinematics*, see below).

In the first part of this chapter I shall further explain the concepts of *activation contour* and *effort*. The purpose of the second part is to clarify what is meant by *dynamics* and *kinematics*, and the way the concepts may be applied to music and movement. The starting point for this is the way the concepts are defined and understood in the field of physics (mechanics), as the distinction here is relatively clear-cut: *kinematics* refers to the movement itself, in the way it may be observed in the way objects move in space, whereas *dynamics* refers to the way forces are applied over time - the cause of the movement. This conceptual demarcation has been applied by research fields concentrating on the movement of the human body, such as biomechanics (Winter, 2005) and human motor control (Rosenbaum, 1991), as well as the line of research concerning the visual perception of biological motion initiated by Johansson (Johansson, 1973). However, when we try to analyse body movement, for example in dance movements, this distinction becomes blurred. Before our eyes we see trajectories in space, the kinematics. The dynamics of the movement, however, are not directly observable; they have to be inferred by observing the kinematics. It appears that when observing human-body movement, dynamics and kinematics are intertwined, so that they are not easy to separate in analytical work.

When we try to apply the terms to music theory yet another issue arises. The kinematics of an event implies that we have an object with a mass that is moving in space. So, the question is: in what sense can music be viewed as a moving object, and in what kind of space is this object moving? These questions concern the broader discussion of music and movement that was briefly presented in the preceding chapter; although we have an intuitive idea of “music as motion”, it may be difficult to see what is actually moving in music.

4.2. Daniel Stern’s theory of *vitality affects* and *activation contour*

Daniel Stern’s *The Interpersonal World of the Infant* (Stern, 2000) is primarily a contribution to the understanding of the way small children, from birth to the phase of language acquisition, develop in an interplay with other people. He presents a theory about the development of early communication skills and the emergence of a sense of self that is grounded in the way the infant relates to its mother/father. The point of departure is Stern’s core interest in the infant’s subjective experience of a self, and the pre-verbal nature of this subjective experience. He proposes that a sense of self develops through four phases:

> Four different senses of the self will be described, each one defining a different domain of self-experience and social relatedness. They are the sense of an *emergent self*, which forms from birth to age two months, the sense of a *core self*, which forms between the ages of two and six months, the sense of a *subjective self*, which forms between seven to fifteen months, and a sense of a *verbal self*, which forms after that. These senses of self are not viewed as successive phases that replace one another. Once formed, each sense of self remains fully functioning and active throughout life. All continue to grow and coexist (Stern, 2000; p 11).

In my review of Stern’s theory, the part concerning the formation of a self will not be discussed any further. I shall concentrate on aspects of the theory that describe the way the caregiver and baby are sensitive to cross-modal correspondences. These correspondences are based on pre-verbal utterances in the form of dynamical shapes, which in Stern’s theory are referred to as *activation contours*.

Initially, the concept of an activation contour refers to the way an event or an action is characterised by a dynamical profile or changes in intensity over time. Second, an activation contour, as an underlying feature of experienced *vitality affects* (see below), is emphasised as an important ingredient of non-verbal communication. Thus, in addition to the concept of an activation contour, what I would see as relevant to my investigation of music-movement relations is the way Stern’s work may be viewed as an attempt to work out a theory that describes how pre-verbal experience is organised, at this pre-verbal stage in development without being constrained by symbolic thought.

> [...] it must be asked, what kind of a sense of self might exist in a preverbal infant? By ”sense” I mean simple (non-self-reflexive) awareness. We are speaking at the level of direct experience, not concept. By ”of self” I mean an invariant pattern of awarenesses that arise only on the occasion of the infant’s actions or mental processes. An invariant pattern of awareness is a form of organization. It is the organizing subjective experience of whatever it is that will later be verbally referenced as the ”self”. This organizing subjective experience is the preverbal, existential counterpart of the objectifiable, self-reflective, verbalizable self (Stern, 2000; p 7).

As seen from this excerpt, Stern presupposes that pre-verbal experience is organised as some kind of invariant patterns of awareness, i.e. patterns that are stored in the memory and recognised from one moment to another on the basis of their direct dynamical/rhythmical features, and not grounded in an interpretation of symbolic/semantic/verbalisable meaning. It is important to note that Stern assumes that the pre-verbal mode of experience is not exclusive to small children. As pointed out in the citation above, each sense of self continues to develop in parallel throughout life, suggesting that an organisation of pre-verbal experience continues to play a prominent role even after childhood. This means that non-symbolic thought is understood as a mode of experience that co-exists and interacts with symbolic thought.

This view is further elaborated on in the more recent contribution *The Present Moment in Psychotherapy and Everyday Life* (Stern, 2004). From the perspective of a psychiatrist, he describes a therapeutical approach that takes its point of departure in brief, everyday moments and he explores these short time-spans of 3 – 10 seconds in detail. It is an approach that dwells on the feelings, the imagery and the thoughts that are evoked by quite mundane, everyday experiences, such as opening the fridge in the morning and realising that there is no butter for breakfast. Stern demonstrates through a number of examples that these *present moments* have the potential to disclose important aspects of our lives. In the therapeutic setting they are explored in terms of their preverbal aspects, i.e. with respect to *how* feelings, imagery and thoughts are evoked. When opening the fridge and seeing no butter, exactly *how* does the feeling of disappointment emerge? Or: *how* do sensations spread through the body; do they explode or do they spread more gradually; do the sensations induce actions, and *how* can these actions be described in terms of intensity and energy; and what kinds of thoughts come to the mind, and *how* do the thoughts emerge, gradually or abruptly?

The *howness* of experience is explored with respect to dynamical, kinematical and rhythmical features. As suggested in the title of Stern's book, such brief present moments are not meaningful just in the therapeutic setting; the dynamical, kinematical and rhythmical howness of subjective experiences may be understood as an ever-present stream of non-verbal consciousness that constitutes the sense of self and that co-exists and co-evolves with symbolic thought in everyday life. This means that we know ourselves through a blend of non-verbal responses/awarenesses and verbal/symbolic thought. Likewise, the way we articulate ourselves and experience other people's articulations is characterised by this blend of non-verbal and symbolic aspects of meaning, i.e. how these different aspects of meaning constrain each other. This theme will be broached again in chapter 5 in a discussion of non-symbolic and symbolic aspects of meaning in music and movement. Stern's contribution in this regard is to shed light on the significance of subjective, non-verbal experience; the way it is vividly meaningful, although not in the symbolic, verbalised sense.

4.2.1. The sense of *howness*: activation contour and vitality affects

The key to this sense of *howness* is the concept of an *activation contour*; the dynamical profile of an emotion or an action. It is closely related to the notion of *vitality affects*. To explain the concept of an activation contour, and the way it is connected to the expression of vitality affects, I shall return to the experiential context that is Stern's point of departure, i.e. the context of the caregiver/pre-verbal infant.

The world of the infant may be described as a world of persons, emotions, objects and events, which are not yet named and not clearly distinguished from each other. Despite the lack of verbal language, the infant is engaging in close, rich and meaningful interactions with its caregivers. Early communication is based solely on actions, sounds, imitations and the expression of emotions. The infant and the caregiver respond reciprocally and with great sensitivity to the rhythmical/dynamical features of actions and emotional expressions. The interplay takes place on a pre-verbal level, which implies that the semantic or symbolic meanings of objects and events are not yet important; the interaction reflects the play itself and the way the interaction establishes and maintains the relation between mother and infant.

One pre-requisite for this interaction to occur is the capacity to perceive sensations to different sensory modalities as being similar, i.e. *amodal perception*. As explained in chapter 3, a set of perceptual attributes are understood as *amodal*, i.e. as attributes that are perceived across the sensory modalities (e.g. intensity, rhythm, shape and duration). At the time Stern published his book *The Interpersonal World of the Infant* this capacity had also been experimentally demonstrated in infants (Lewkowicz & Turkewitz, 1980; Meltzoff & Borton, 1979), and Stern builds on this research. Within the caregiver/infant context, amodal perception plays a central role as both the mother and the baby acknowledge that a sound may be heard as being similar to a seen movement, for example the mother tunes in to the sounds of her baby by nodding her head or making hand movements in a similar rhythm.

Stern illustrates the amodal/crossmodal character of such interactions by referring to examples, e.g.:

> A nine-month old boy bangs his hand on a soft toy, at first in some anger but gradually with pleasure, exuberance and humor. He sets up a steady rhythm. Mother falls into his rhythm and says, ”kaaaaa-*bam*, kaaaaa-*bam*,” the ”*bam*” falling on the stroke and the ”kaaaaa” riding with the preparatory upswing and the suspenseful holding of his arm aloft before it falls (Stern, 2000; p 140).

Stern argues that categories of emotion, such as joy, anger, pleasure, are too broad and general to understand this kind of communication. Hence, he introduces the concept of *vitality affects* to denote the almost infinite variations of an emotion, for example joy. A smile can be explosive or it can emerge more gradually, or in degrees between the sudden and the gradual. The same applies to actions; any action, such as closing a door, can be performed with a wide range of shadings.

> There is a [...] quality of experience that can arise directly from encounters with people, a quality that involves vitality affects. What do we mean by this, and why is it necessary to add a new term for certain forms of human experience? It is necessary because many qualities of feeling that occur do not fit into our existing lexicon or taxonomy of affects. These elusive qualities are better captured by dynamic, kinetic terms, such as ”surging”, ”fading away”, ”fleeting”, ”explosive”, ”crescendo”, decrescendo”, ”bursting”, ”drawn out”, and so on. These qualities are most certainly sensible to infants and of great daily, even momentary, importance (Stern, 2000; p 54).

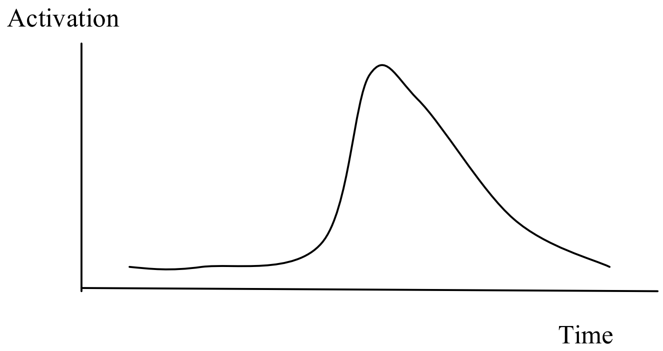
Each of these variations can be described in terms of an *activation contour*, the dynamic profile; the way the dynamics, or the intensity, of the emotion or the action evolve in time. An activation contour is understood as an underlying feature of vitality affects, meaning that the concept of an activation contour is introduced by Stern to capture the unique performance of any emotion or expressive action. The action characterised as ‘slamming’ the door is clearly distinguished from closing the door with a ‘gentle push’. And between the two extremes of ‘slamming’ and ‘gently pushing’ there are an infinite number of variations to which we are extremely sensitive, and which make a difference in communicative relationships. These variants are characterised by patterns of changes in activation, i.e. activation contours[20].

The concept of activation is understood as an amodal feature. This means that a specific vitality affect may be articulated across modalities based on the underlying feature of activation as an intermodal form.

> Because activation contours (such as “rushes” of thought, feeling, or action) can apply to any kind of behavior or sentience, an activation contour can be abstracted from one kind of behavior and can exist in some amodal form so that it can apply to another kind of overt behavior or mental process. These abstract representations then permit intermodal correspondences to be made between similar activation contours expressed in diverse behavioral manifestations (Stern, 2000; p 57).

In the example of a boy banging on a toy and the mother joining in by performing a rhythm with her voice, there is a certain element of imitation that is based on this amodal nature of activation contours. We can imagine how the child is lifting his hand and hits the toy with a quick, firm stroke, and without pausing, repeats this again and again in a regular rhythm. One phase of this recurrent pattern may be described by the activation contour; a steeply increasing intensity in the preparatory phase (falling hand), an energetic stroke (hits the toy), and a quick release of energy while building up energy for a new stroke (rising hand). The movements are characterised by an abruptness, a highly energetic rhythm. The mother captures the character of this rhythm and synchronises the accent, the *bam*, with the stroke. We can also imagine from the description that the pitch characteristics of the ”kaaaa”, and its crescendoing feature, resemble the intensity profile of the falling hand-movement, so that the overall activation contours of the boy’s movement pattern and the mother’s vocal pattern are similar. Figure 9 illustrates schematically the way we may imagine that the level of activation varies in time, the shape of the curve visualising the overall activation contour:

[20] In a footnote Stern comments that his notion of the activation contour as an underlying feature of vitality affects builds on the work of Tompkins. However, whereas Tompkins concluded that discrete patterns of neural firing (density x time) result in Darwinian categories of affect (such as joy, pleasure etc.) Stern argues that patterns of neural firing correspond to an experiencing of vitality affects. Thus, “all the different activation contours may be described in terms of intensity or sensation as a function of time. Changes in intensity are adequate to explain “explodings”, “fadings”, “rushes”, and so on, no matter what actual behavior or neural system that is the source of these changes” (Stern, 2000; p 57).



Boy: Holds hand high –falls - hits - lifts – (holds hand)
Mother: ”kaaaaaaaaaaaaaaaa” - ”*bam*” - breathes in

Figure 9. The ascending-descending contour of the curve suggests an increasing-decreasing contour in activation level (or change in intensity over time). The activation contour is similar for both the gesture of the boy and the singing of the mother, so that they are matched amodally. When the cycle is repeated a temporal pattern, or rhythm, emerges. It should be noted that the graph does not depict the shape of the hand’s trajectory.

4.2.3. Cross-modal correspondence: imitation and affect attunement

The example describes an instance of audio-visual correspondence, and the way this correspondence is fundamental to the interplay of mother and child: the mother captures rhythmic qualities of her child’s activity; there is a match between the *howness* of the banging and the *howness* of the singing. This cross-modal correspondence of vocalization and sound is described in terms of dynamical aspects (changes in intensity). The changes in dynamics emerge as a recurrent temporal pattern of contrasts between stronger and weaker intensities. Although not explicitly pointed out, it seems that Stern uses the expression *temporal pattern* and the term *rhythm* interchangeably so that both refer to processes characterised by a pattern of changes in dynamics (Stern 2000; pp 51 and 53).

The nature of this correspondence, the way Stern observes and discusses the interplay of simple cross-modal imitation and a shared experience of “inner feeling state”, is a core theme in Stern’s theory. Initially, the communication is based on imitation: the mother’s imitation of her son’s activity is a pre-requisite for establishing and maintaining an interaction. This is a kind of communication that prevails throughout the first months of the baby’s life; however, from a certain stage a new dimension is added. The mother introduces what Stern denotes as *affect attunement*, i.e. she engages in the play in such a way that the child experiences that the mother “knows” the inner-feeling state of the child. From this stage, pure imitation is not sufficient for the child to develop further. In Stern’s terms, *imitation* is understood as a reproduction of overt behaviour, whereas *affect attunement* refers to features in the mother’s execution of actions, i.e. subtleties in the mother’s performance that not only reflect or mimic the overt behaviour of the child, but that are rooted in the mother’s intention of making known to the child that she understands the inner-feeling state of the child.

Stern discusses the way imitation and affect attunement are inter-related. In the following quote, he lists three important aspects of affect attunement:

1. They give the impression that a kind of imitation has occurred. There is no faithful rendering of the infant's overt behavior, but some form of matching is going on.
2. The matching is largely cross-modal. That is, the channel of modality of expression used by the mother to match the infant's behavior is different from the channel or modality used by the infant. [...]
3. What is being matched is not the other person's behavior *per se*, but rather some aspect of the behavior that reflects the person's inner state. The ultimate reference for the match appears to be the feeling state (inferred or directly apprehended), not the external behavioral event. Thus the match appears to occur between the expressions of inner state. [...] (Stern, 2000; p 142)

Although affect attunement implies a quality of imitation, affect attunement and imitation are distinguished from each other by their different focuses of attention; whereas imitation directs the focus to the overt behaviour, attunement places the focus on the inner feeling of an action. On the other hand, it appears that Stern does not want to place imitation and attunement in two separate categories; he acknowledges them as two closely inter-related processes.

Affect attunement, then, is the performance of behaviors that express the quality of feeling of a shared affect state without imitating the exact behavioral expression of the inner state. If we could demonstrate subjective affect-sharing only with true imitations, we would be limited to flurries of rampant imitation. Our affectively responsive behavior would look ludicrous, maybe even robot-like.

The reason attunement behaviors are so important as separate phenomena is that true imitation does not permit the partners to refer to the internal state. It maintains the focus of attention upon the forms of the external behaviors. Attunement behaviors, on the other hand, recast the event and shift the focus of attention to what is behind the behavior, to the quality of feeling that is being shared. [...] In actuality, however, there does not appear to be a true dichotomy between attunement and imitation; rather, they seem to occupy two ends of a spectrum (Stern, 2000; p 142).

On the one hand, Stern states that attunement is distinguished from imitation on the basis of the kind of intuition, or empathy, that most adults exhibit when with small children. On the other hand, since attunement involves an element of imitation he tries to identify the aspects of behaviour that "could be matched without actually imitating them" (Stern, 2000; 146)[21]. Based on observations of mother/infant interactions, he proposes that three features may be important, all considered to be amodally transferable. These are intensity, timing and shape. He then lists a set of features that in more detail explains how these overall features may be involved in affect attunement:

1. *Absolute intensity*. The level of the intensity of the mother's behavior is the same as that of the infant's, irrespective of the mode or form of the behavior. [...]
2. *Intensity contour*. The changes of intensity over time are matched. [...]
3. *Temporal beat*. A regular pulsation in time is matched. [...]

[21] In this quote it appears that Stern uses the term *match* for an approximate correspondence, whereas *imitation* refers to a robot-like, exact rendering.

4. *Rhythm.* A pattern of pulsations of unequal stress is matched. [...]
5. *Duration.* The time span of the behavior is matched. [...]
6. *Shape.* Some spatial feature of a behavior that can be abstracted and rendered in a different act is matched. [...] (Stern, 2000; p 146).

To sum up this section, the concept of the activation contour is closely related to the experiencing and expression of vitality affects. Characterised by dynamical and rhythmical features, i.e. the way fluctuations in intensity are shaped and patterned in time, activation contours are understood as meaningful in the pre-verbal/non-symbolic sense. This meaningfulness is contained in the way shadings of activation contours describe the howness of an emotion or action. Finally, the intermodal character of activation contours may be observed within the caregiver/infant context, i.e. the way this non-verbal communication utilises the sensitivity to cross-modal correspondences as a resource.

Questions related to children's imitative behaviours are central to Stern's observations. In this regard, it should be noted that a more recent contribution discusses the accuracy of *imitation*; i.e. when we try to imitate the movements of another person, how 'good' is the match required to be to be accepted as an *imitation*? Studying imitative behaviour experimentally, Wohlschläger and colleagues have proposed that imitation is characterised by the way we try to figure out the target points, or certain salient or accentuated moments of an event, and that the level of precision is higher around such accentuated moments, whereas the match may be less precise before and after these moments (Wohlschläger, Gattis, & Bekkering, 2003). This would suggest that an action that is intended to imitate another action does not have to be a perfect reproduction to be regarded as an *imitation*. This will be commented on further in chapter 5.

4.3. The concept of *effort* in Laban Motion Analysis

Laban, a choreographer and dance teacher, developed a conceptual framework as a method for the observation and analysis of dance, known as Laban Movement Analysis (LMA). The underlying theoretical considerations and analytical approach are outlined in Laban's major works, *Modern Educational Dance* (Laban, 1948), *Mastery of Movement* (Laban, 1971)[22], *Choreutics* (Laban & Ullmann, 1966), and *Effort* (Laban & Lawrence, 1947). As a descriptive approach LMA includes four perspectives, i.e. four angles that interact in human movement and which in analysis may be approached one at a time: *body* (how is the body organised, for example in terms of connectivity), *shape* (what forms does the body make, and how does the shape change?), *space* (how does the mover relate to space?), and finally, *effort* (Hackney, 2002). In this review I shall concentrate on the *effort* aspect of the approach.

Laban introduced and defined *effort* as a theoretical concept to distinguish between living movement and mechanical movement, i.e. effort is a feature found

[22] This refers to the third edition. The first edition was published in 1950 under the title *Mastery of Movement on the Stage.*

exclusively in human-body movement[23]. For example, this means that effort does not characterise the movement of a car or any other inanimate object. Thus, the understanding of the term effort in the context of Laban's theory goes beyond the restricted dictionary meaning, e.g. the physical effort required to make a movement. According to Laban, effort is the origin of any human bodily movement; effort is the inner quality or source of the movement (Laban, 1971). On the one hand, effort includes physical aspects of the movement, i.e. the way forces are involved in producing the movement; the way these forces are distributed in time. On the other hand, the concept of effort tries to capture the mental attitude of the movement, such as intentional and motivational aspects.

Since effort seems to reside in a two-sidedness, between the physicality and the intentionality of the movement, effort is not easily grasped and defined as a theoretical concept. I think the best way to aid interpretation of the concept is to learn how a community of users, e.g. dancers and choreographers, describe the concept in practical work[24] and in writings. Thus, I shall first quote the way Laban himself explained the term, including the notion that effort may be heard in sound:

> In order to discern the mechanics of motion within living movement in which purposeful control of the physical happening is at work, it is useful to give a name to the inner function originating such movement. The word used here for this purpose is *effort*. Every human movement is indissolubly linked with an effort, which is, indeed, its origin and inner aspect. Effort and its resulting action may be both unconscious and involuntary, but they are always present in any bodily movement; otherwise they could not be perceived by others, or become effectual in the external surroundings of the moving person. Effort is visible in the action movement of a worker, or a dancer, and it is audible in song and speech. If one hears a laugh or a cry of despair, one can visualise the movement accompanying the effort (Laban, 1971; p 24).

The understanding of effort as a feature of movement that includes both physicality and mental processes is articulated by one of Laban's successors, Betty Redfern (Redfern, 1965):

> […] the performance of sequence composed of certain effort elements and shape structures can produce those moods and feelings which correspond to them: movement involves change not only of a physical character but mental and emotional change too. […] The art of movement, therefore, involves more than the mobilising of joints and strengthening of muscles, for it is concerned with the study of reciprocal influences of bodily action and mental and emotional processes (Redfern, 1965; pp 5-6).

Moore and Yamamoto (Moore & Yamamoto, 1988) further interpret the way effort is connected to the intentionality of movement:

> The human being moves to satisfy a need. Actions are guided and purposeful, and the intentions of the mover are made clear by the way in which the person moves. While the uses of space and of the body reveal the mover's purposes, Laban believed that the uses of energy, or the dynamics of

[23] The term *effort* will throughout the thesis be used in accordance with the framework of understanding provided by Laban Motion Analysis.
[24] In this regard I am indebted to the choreographer Andre Austvold who gave me my first practical introduction to Laban Motion Analysis.

an action, were particularly evocative of intentions. Thus, he used the term, “effort”, to delineate the dynamic energies exerted in movement [...] (Moore & Yamamoto; p 185).

As a concept underlying movement analysis, effort is further elaborated on in a set of analytical/descriptive terms, i.e. what are known as the motion factors of *Time*, *Weight*, *Space* and *Flow*. The terms are used to denote categories on which observation and analysis are based. They are derived from the core concept of effort, which means that a movement, e.g. according to the *Time* factor, may be characterised along a continuum between *sustained/slow* or *quick/sudden* effort. Similarly, a movement may be described in terms of the *Weight* factor, i.e. as being performed with a *firm/heavy* effort as opposed to a more *gentle/lighter* effort, or degrees between the two extremes. Table 1 provides an overview of the motion factors and their corresponding *effort elements*:

Effort element	Firm: strong/heavy, feeling of weightiness	Sudden: quick speed, feeling of short span of time	Direct: straight line, feeling of narrowness	Free: not controlled
	↑	↑	↑	↑
Motion factor	**Weight**	**Time**	**Space**	**Flow**
	↓	↓	↓	↓
Effort element	Gentle: weak/light, feeling of weightlessness	Sustained: slow speed, feeling of a long span of time	Flexible/indirect: wavy line, feeling of everywhereness	Bound: hampered or controlled

Table 1. Overview of motion factors and their corresponding *effort* elements.

The following will further elaborate on the way each effort element is understood as a phenomenal quality. The descriptions are based on the writings of Laban himself (Laban, 1971; Laban & Lawrence, 1974), as well as on more recent contributions about Laban theory (Hackney, 2002; Moore & Yamamoto, 1988; Redfern, 1965).

Weight. A movement can be performed in a gentle or a firm manner, inducing a feeling of weightlessness or weightiness. A heavy/firm *Weight* is associated with the feeling of letting the weight of the body push downwards and into the ground, the feeling of heaviness; whereas a light *Weight* is characterised as the feeling of being weightless; the feeling of stretching upwards as though gravity is defied. Movements performed with a heavy/firm effort are characterised as powerful, forceful, having a firm touch and impactful, whereas a light/gentle effort is characterised as airy, delicate, having a fine touch and as being buoyant (Hackney, 2002; p 220). It should be noted that the *Weight* factor has also been termed *Pressure* (Moore & Yamamoto, 1988).

Time. A movement can be quick or sustained/slow. Hackney associates the sustained *effort* with movements that are performed in a leisurely, gradual, lingering and prolonging manner, whereas the sudden effort is found in urgent, quick, instantaneous and staccato movements (Hackney, 2002; p 220). The *Time* factor does not restrict itself to an evaluation of displacement in time as in a quick motion; analysis of effort does also include a further observation of the quality of quickness. Two movements may be similar to each other in terms of velocity, but they may differ in what we might term "mental attitude", e.g. one of the movements is "felt" as hurried, whereas the other travels along the same trajectory in a more relaxed manner.

Space. This motion factor refers to the focus of a moment, and has thus in some explanations of the theory been termed *Focus* (Moore & Yamamoto, 1988). For example, putting the key in a lock and turning it has a clear, direct focus. On the other hand, waving to a crowd of people may have a flexible, indirect quality, leaving a sense of everywhereness. In Hackney's interpretation of the *Space* factor, the indirect effort opposes the direct effort as follows: multi-focussed vs single-focused; flexible attention vs channelled; all-around awareness vs pinpointed; and finally, all-encompassing vs laser-like (Hackney, 2002; p 221).

Flow. The last factor, the quality of *Flow*, refers to the degree of control involved in the movement. Binding *Flow*, or controlled effort, is seen in actions that are performed so that they are ready to stop any time, i.e. the movement trajectory of for example an arm may be stopped and re-directed at any moment. This effort element is found in controlled and precise movements, such as in "carrying a cup brimming with hot coffee" (Moore & Yamamoto; p 201) and are characterised by being controlled, careful, contained and restrained (Hackney, 2002; p 219). Free *Flow*, on the other hand, is characterised by the feeling of letting go; such as when the body, or a body part, is set in motion with an impulse, and then keeps on moving wherever it 'wishes'. Thus, free *Flow* is characterised as being an outpouring, fluid, a release and liquid (Hackney, 2002), as well as abandoned, fluent, easy-going and unrestricted (Moore & Yamamoto, 1988).

The basic idea of the approach is that each movement can be described as combinations of these motion factors, in terms of their effort elements. Laban outlines eight different combinations of *Weight*, *Space* and *Time* that he considers to be *basic effort actions*. The overall qualities of these actions are characterised as: 'slashing', 'gliding', 'pressing', 'flicking', 'wringing', 'dabbing', 'punching' and 'floating'. For example, the action of 'slashing' is performed with a heavy *Weight*, indirect *Space* and quick *Time*, whereas the contrasting 'gliding' is characterised by a light *Weight*, direct *Space* and sustained *Time* (Laban, 1971; pp 79-80). Approaching analysis by understanding a movement as a combination of co-occurring effort elements, or aspects of movement, is a basic understanding of LMA.

A further underlying notion of Laban theory is that movement processes are fundamentally characterised by the way effort elements change over time (Moore & Yamamoto, 2002; p 183). The main objective of analysis is to characterise such changes. The analysis is not limited to evaluating whether the movement is performed either with a gentle or a firm effort, or something in between, but an important feature of the analysis is to examine the way the temporal unfolding of the movement can be characterised by changes in effort, i.e. changes from gentle to firm, as well as whether the changes are more or less abrupt or more or less gradual. The arrows in table 1 illustrate this point. They are meant to indicate that the underlying effort may fluctuate between degrees of gentleness and firmness, and that the effort elements do not denote static features. This suggests that the effort elements might be more properly expressed in the terms gentler and firmer. An analysis of each of the four motion factors are meant to reveal such changes in effort, and these changes are thought to contribute to the rhythmical aspect of the movement.

As is apparent in this review of *activation* and *effort*, the concepts have obvious overlapping aspects; first, in the way they focus on dynamical aspects, and second, in the sense that concepts seem to be introduced on the basis of a special interest in the way change characterises a process.

Regarding the applicability of the concepts to the experiencing of sonic processes, Stern's activation is basically amodal in nature. And as mentioned in the introduction to the chapter, by virtue of Laban's idea that effort may be seen in a movement and by imagination sensed in a sonic utterance, the possibility of applying effort to analysis of musical performance seems plausible. I will come back to this.

4.4. Dynamics and kinematics in music and movement

To illustrate how music and movement correspondences may be understood in terms of the kinematics-dynamics, I shall first return to the examples of music-movement relations that were introduced in the first chapter of this thesis: on the one hand the contrast between Cunningham's great, forceful ballet leap and the 'thin', high-pitched violin tone; on the other hand, the direct, 'mickey-moused' correspondences found in cartoons. Regarding the Cunningham example, I commented on this in chapter 2, and suggested the following analysis:

- The great ballet leap was performed with a strong and abruptly applied force, whereas we may imagine that the violin tone was produced with a weaker and continuously applied force, i.e. we have a mismatch in terms of *dynamical aspects.*
- The leap is also characterised by a high trajectory in space; similarly, the violin tone is high-pitched, i.e. we have a match in terms of *kinematical aspects*.

This last point implies that kinematics in music is understood as being linked to pitch characteristics. This will be further discussed in the following sections. With this understanding in mind we may analyse 'mickey-moused' relations as follows:

- The heavy tread of a massive body corresponds to big, heavy chords in the music, i.e. a *dynamical* match

- The light steps of a small body corresponds to high pitched pizzicato melodic patterns in the music, i.e. a *dynamical* match
- A character running upstairs corresponds to an upward, scale-like, melodic contour, i.e. a *kinematical* match
- Finally, if we acknowledge that kinematics in music are also connected to changes in speed (see discussion later in this section), a running character slowing down to a halt may correspond to a deceleration (ritardando) of a similar rate in the music, i.e. a *kinematical* match

It seems that music and movement in these examples correspond to one other with regards to kinematical aspects and/or dynamical aspects, and the way these aspects in movement and music respectively are brought together by an adequate degree of synchronisation, meaning that 'something' in the movement that is similar to 'something' in the music occurs simultaneously.

4.4.1. *Dynamics* and *kinematics* in movement

Originating from the field of mechanics, and applied to research fields concerning human motion such as biomechanics, motor control theory and the visual perception of biological motion, the kinematics of a movement is commonly defined as follows:

> Kinematic variables are involved in the description of the movement, independent of the forces that cause that movement. They include linear and angular displacements, velocities and accelerations (Winter, 2005; p 9).

As already mentioned, the great leap performed by Cunningham may be described in kinematical terms, i.e. in terms of the trajectory in space and the pattern of the change in velocity along this trajectory. The following captures the difference between kinematics and dynamics:

> The science of mechanics distinguishes kinematics, motion described as such, from dynamics, which is motion explained in terms of what causes and constrains it (Runeson & Frykholm, 1983); p 587).

Cunningham's great leap is characterised by the forceful impulse that initiates the leap into the air. The leap resembles a *ballistic* type of movement, which involves an object being thrown into the air with one initiating force, and then left alone without being influenced by any new forces apart from gravity and air resistance (friction). A contrasting kind of dynamical pattern would be when force is applied more *continuously*, for example when lifting an object from the ground or shoving a heavy object forward.

A further major difference between dynamics and kinematics is that the kinematics are directly observable, whereas the dynamics are hidden from our eyes as they take place in the coordinated, muscular activity inside the other person. This means that the dynamics have to be inferred based on what we see, i.e. the kinematics. In fact, it has been proposed that recognizing human-movement patterns is based on the inferred dynamical pattern, which is specified by the kinematics.

> Using the kinematics versus dynamics distinction in this way we have previously challenged the traditional approach to motion perception by suggesting that perception of events is predominantly in terms of dynamic rather than kinematic properties. Put simply, we tend to perceive causal aspects of events, not movements as such (Runeson & Frykholm, 1983; p 588).

This proposal would imply that the ballistic quality of Cunningham's great leap is the salient feature of the event; that the dynamic pattern fundamentally characterises the perceptual feature of the leap. The proposal should be understood in the context of a line of research into the visual perception of biological motion which was initiated by Johansson (Johansson, 1973). Johansson's major contribution was to use sparse point-light displays to study the perception of human motion. Subjects were asked to recognize patterns of movement as represented by moving point-lights placed on the joints of the actor and displayed on an otherwise dark screen. The research demonstrated how sensitive we are to human movement patterns. Over the years, different explanations have been offered to account for the phenomenon (see Shipley, 2003 and Pollick, 2004 for reviews). The *kinematic-specification-of-dynamics* hypothesis proposed by Runesson & Frykholm is one attempt to account for the phenomenon. Runesson & Frykholm ground the kinematic-specification-of-dynamics hypothesis in ecological theory; first, by referring to the notion that perception entails picking up useful information, which implies that perception is relatively direct[25]; and second, the assumption that dynamical features play a prominent role in event perception (Runesson & Frykholm, pp 586-588).

In a more recent contribution, Shipley (Shipley, 2003) suggests that the explanations for the immediate nature of human-movement recognition "fall into two classes: *event-from-form* and *event-from-dynamics* theories" (Shipley, 2003; p 377). The first implies that the form of the moving object is extracted from the moving point-lights, i.e. that virtual lines are drawn by mental processing between the dots, and that the form is recognised on the basis of such processing. The action is judged and recognised on the basis of this imagined form. Regarding the event-from-dynamics model, I would suggest that the understanding proposed by Runeson & Frykholm would fall into this category, since they assign the pickup of dynamical features a central role in visual perception of movement.

In a review of the accumulated research on the perception of biological motion, Pollick (Pollick, 2004) concludes that although some experiments support the event-from-dynamics hypothesis, uncertainty regarding the underlying cognitive mechanisms still exists.

> Although convincing demonstrations were provided that showed how properties such as lifted weight were judged correctly, an elaboration of the specific mechanisms behind this capability was not provided [...] to date there is still no thorough and convincing theoretical construct to explain the perception of biological motion (Pollick, 2004; p 11).

The mechanisms that Pollick asks for may be found in the proposal suggested by Wilson & Knöblich (Wilson & Knöblich, 2005) that was mentioned in chapter 3.

[25] This refers to Gibson's notion of the direct perception of affordances (Gibson, 1986), see chapter 3.

Based on neurological research and the role of mirror neurons, they suggest that recognizing biological motion is grounded in the covert imitation of other people's actions. What I would suggest to call an *event-from-imitation* hypothesis proposes that while observing others, we continuously imitate their movements "in our minds" and in this manner we explore and predict their actions. The mental imitation of motor actions appears to be an automatic, on-going process, thus serving as a constantly available resource with which to establish a relation to movements in the environment. It should be noted that overt imitation is usually inhibited and hence not overtly executed. This proposal would imply that our access to other's actions is quite direct, as Runeson & Frykholm suggest. However, the event-from-imitation process does not necessarily mean that the dynamics of the observed movement are a more salient feature, i.e. that the dynamics are more readily picked up than the kinematics.

4.4.2. *Dynamics* and *kinematics* in music

For the present I shall leave this discussion of dynamics and kinematics in human movement and turn to music, specifically the role of kinematics and dynamics in music. The question is: does it make sense to use these terms in relation to music, considering that in sound there is no movement in the physical sense? More generally, when talking about music we do make allusions to movement, we talk about music using terms that are usually used to describe movements, but what is moving? What kind of 'object' are we referring to when we talk about movement in music? In what kind of space does the movement take place?

Todd (Todd, 1995) has proposed that kinematics in music may be understood in terms of tempo variations (velocity) and pitch intervals (the distance/length between pitches).

> What, then, are the kinematic variables of musical performance? [...], we may consider a score to be like a trajectory in 2-D space. The vertical axis describes a 1-D *pitch space p* while the horizontal axis describes a spacelike dimension, measured in units of beats or bars, which we refer to here as *metrical position x*. [...] Thus as discussed above, we may discern two distinct kinds of motion in music, *tonal motion*, i.e., pitch as a function of time *p(t)*, and *rhythmic motion*, i.e., metrical position as a function of time *x(t)* (Todd, 1995; p 1941).

Intuitively this does make sense. Most musicians would have a sense of moving a greater distance when going from C3 to G3 than from C3 to E3.[26] Similarly, we talk about the beginning and the end of a piece, as though time resembled a path in the horizontal plane, so that playing the piece is like walking or running at different tempi along this path. In this perspective, the sense of movement in music is metaphorical; experience in the domain of actual, physical movement is transferred to experiencing a sonic phenomenon, which does not include actual physical movement.

In musical discourse the term *dynamics* is commonly referred to in terms of *forte* and *piano*, *crescendo* and *decrescendo*, and degrees of these. Understood in this manner, dynamics appears to describe the perceived effect of loudness or intensity, rather than the actual force that produces the effect. In a paper published before the

---

[26] Midi notation

one focusing on kinematics, Todd discusses a model of expressivity in music which includes a coupling of dynamics and kinematics, thus applying the concept of the terms from mechanics as discussed above (Todd, 1992). As a starting point he refers to the general observation of some particular styles of music, in which the phrasing seems to be quite consistently constituted by a correspondence between tempo and dynamics, i.e. the phrase structure is constituted by an increase-decrease contour in tempo (kinematical variable) and a corresponding increase-decrease contour in musical dynamics (dynamical variable). The musical phenomenon he is referring to is the style of phrasing in Romantic music, the way the musician tends to 'push forward' with increasing tension towards a goal, and then releases the tension towards the end of the phrase. Furthermore, Todd suggests that these typical, simultaneous acceleration/deceleration and crescendo/decrescendo contours in a musical phrase are similar to corresponding kinematical and dynamical patterns of simple motor actions.

In a procedure that he denotes as being a method of analysis/synthesis/evaluation, Todd attempts to appraise the link between perceived expressivity in music and the properties of physical motion. First, based on the principles of mechanics he formulates an initial, mathematical model that is designed to predict the relation between the phrase's tempo profile and dynamical profile. In the next step, performances of a Chopin prelude[27] are recorded and analysed with estimations of tempo (beats per second) and intensity as functions of beat number as outputs. For the dynamical variable, intensity is used as an indicator, and is measured on the basis of hammer flight time, using the piano system developed by Shaffer (Shaffer, 1981). This implies that the dynamics are directly linked to and measured by the forces involved in pressing the keys of the piano, which seems reasonable and in accordance with the standard physical definition of dynamics.

Having made these preparations, Todd uses the initial mathematical models to perform a series of simulations of the tempo-dynamics relation. The model is successively modified by adjusting parameters in order to resemble the performance data. The evaluation is performed using regression analysis, so that the best match between model and performance data is obtained when variance is minimized. Through this procedure, Todd arrives at a model, which, based on the statistical analysis, satisfactorily accounts for the observed data. On the basis of this simulation experiment we may conclude that it is possible to work out a mathematical model that predicts the relation between tempo and dynamics, thus substantiating the notion of a close relation between music and physical motion.

In another contribution to the same topic and published in the same year as Todd's paper on dynamics, Feldman and his colleagues (Feldman, Epstein, & Richards, 1992) similarly investigate the relation between kinematics and dynamics. In their paper, however, measuring dynamics in terms of "intensity" is not included. Dynamics are understood as the force applied over time to change the tempo in music, referring to the intuitive understanding of performing musicians, ”that changing the tempo of the musical beat involves something like the application of a force to the beat, as evidenced by such figures of speech as ’pushing forward’, for an accelerando, and ’holding back’, for a ritard" (Feldman et al., 1992; p 187).

[27] Prelude in F sharp minor, Op 28, no. 8

In mechanics, force is rendered by the equation $F=m*a$ where m is the *mass* of the object to which the force is applied, and a is *acceleration.* Acceleration is rendered by the equation $a=v/t$, where v is *velocity* and t is *time*. In Todd's work mentioned above, velocity in music is the tempo, i.e. the metrical distance travelled in time, expressed by $v=\Delta x/t$, where $\Delta x=x2-x1$ is the distance between two moments in time. Accordingly, acceleration is the changes in tempo over time, so that the velocity at any given moment is the derivative of the acceleration. In sum, we have three basic equations: the first expresses the dynamics ($F=m*a$), the second expresses the kinematical variable ($v=\Delta x/t$), and finally the equation $a=v/t$ links dynamics and kinematics.

This begs the question: What is the "mass" in music? To work out a force model for expressivity in music, Feldman and colleagues therefore introduce an abstract notion of a particle with mass, so that this particle's position P is a function of time. The function $P=P(t)$ will inevitably describe a steadily ascending graph, but at a changing rate depending on the tempo fluctuations in the music.

> The progression of music through time can be conceived of as the movement along the time axis of a (purely abstract) "particle of music," whose position at time t is given by monotonically nondecreasing $P=P(t)$. [...] The particle has some "mass" m, which may be thought of naturally as the resistance of the music to change in tempo (depending, we might imagine, on musical factors such as the emotional "heaviness" of the passage, or on performance parameters such as the ease of maintaining ensemble); or else, again, as a purely formal parameter. Throughout the following analysis, it should be kept in mind that the "forces" we are discussing are not physical but rather mental components of some plan of action residing in the mind of the performer (Feldman et al., 1992; p 188).

Based on this presupposition, they suggest the basic hypothesis that tempo changes in music stand in a mathematical, predictable relation to the applied force: first, that a force applied linearly will cause a change in the tempo of the music with a quadratic tempo profile; or alternatively, that a force applied quadratically will cause a change in tempo in music with a cubic tempo profile. These two mathematical models, the *linear force-to-quadratic tempo* and the *quadratic force-to-cubic tempo*, are then evaluated against performance data which are basically tempo profiles of five different recorded music samples. Regression analysis is used to evaluate the match between tempo models and performance profiles. Feldman and colleagues sum up their analyses by suggesting that except for one of the samples the "other four examples and their derivatives all have shapes that are consistent with simple low-order force events, where the forces are either linear or a simple quadratic form, and the resulting tempo is either parabolic or spline-shaped. The force model thus accounts neatly for the observed tempo profiles (Feldman et al., 1992; p 202)."

Although the conclusions rely on the assumption of the existence of mass in music, their findings lead Feldman and colleagues to suggest that the general notion of motional qualities in music, which can be traced in the way we talk about music, is not merely metaphorical. The way the tempo profiles conform to the physical models points in the direction that tempo variations have a fundamentally physical character.

> [...] this conformance seems to reflect an underlying, unconscious conception of music as a quasi-physical thing that ”moves forward” as it unfolds through time, now speeding up and now slowing down, in accord with the moment-to-moment flux in its rhythmic, harmonic, and affective character – a conception reflected in musicians’ common use of terms such as ”movement”, ”motion”, and ”flow” to characterize the progression of music. We propose in effect that the relation between tempo change (designated by the score and manipulated by the performer) and real physical movement (controlled by physical forces) is somewhat more than just a metaphor: the formal machinery is largely the same (Feldman et al., 1992; p 202).

As mentioned in chapter 2, the relation between music and physical movement was also studied by Kronman & Sundberg (Kronman & Sundberg, 1987). The reader may recall that they compared timing properties of retardation in music performance, i.e. in performances referred to as “motor music”, with a model for deceleration of a runner slowing down to a full stop. They found that the average timing profile of the musical ritards matched the model for timing of physical movement well. Studies of this kind suggest a link between timing of music, which is certainly an important aspect of musical expression, and body movement. In light of this, the notion of movement in music appears to be real and more than just a metaphor. On the other hand, to make the basic concepts of movement work in the context of music, dynamics and kinematics, we are forced to make leaps of reasoning; we make assumptions, such as mass in music, the (vertical) distance between two pitches in space, and the (horizontal) distance between beats in a metrical grid. As these assumptions appear to be metaphorically based, and we are still not able to state what is actually moving in music, and in what kind of space, it seems to be safer to uphold the perspective that music is related to motion by metaphorical transfer.

We can see that the discussion of dynamics and kinematics broaches the more general issue of what is moving in music; how do we understand the relation between movement and music? The question, or more specifically, ”what is moving” and ”in what kind of space”, is raised by Eric Clarke in his recent book *Ways of Listening* (Clarke, 2005). He suggests that the ”relationship [between music and motion] is truly perceptual rather than metaphorical, symbolic or analogical (Clarke, 2005; p 74).” Based on an ecological theory of perception and cognition, he proposes that music, as with other everyday, sonic phenomena, should be understood as an event. What characterises an event, he maintains, is that it has a source, and that it is embedded in its surroundings. The source of sounds is movements and the way movements make “things” sound. The role of auditory perception is to detect the event; to establish a relation between the sonic event and the perceiver. Clarke refers to McAdams (McAdams, 1984) and Bregman (Bregman, 1990) and their notion that what distinguishes everyday sounds and music is that, in the first case, the source is real, whereas with music the source is virtual. However, ”the psychological processes involved in perceiving real and virtual sources are identical, just as the experiences themselves may be equally vivid and apparently veridical – but while one refers to real objects and events in the world, the other does not” (Clarke, 2005; p 72). In this lies what Clarke understands as being the true perceptual reality of the relation between music and motion.

If we cease to insist that music and everyday sounds are phenomena residing in separate worlds of experience, and give up the idea that sound is perceived as pure, i.e. separated from the source, this bring us a step closer to an understanding of the relationship between music and motion that is not based merely on metaphorical transfer. This understanding is grounded in acknowledging the role of the source of the auditory event, the intimate link between the sound and the action producing the sound. It is a fact that there are actual physical movements that are directly related to musical sound and these are the sound-producing actions of musicians. If we apply the fundamental notion of ecological theory/motor theory of perception that perception is imagined action (Berthoz, 2000), i.e. that perception of sound involves imagining or simulating the actions that produced the sound, the relation between music perception and motion becomes a perceptual reality. The role of sound-producing actions in music perception and imagery has been pointed out by many authors, e.g. Shove & Repp (Shove & Repp, 1995) and Godøy (Godøy, 2001). Clarke interprets the role of sound-producing actions as follows:

> The basic principle can be stated simply enough: since sounds in the everyday world specify (among other things) the motional characteristics of their source, it is inevitable that musical sounds will also specify movements and gestures – both the real movements and gestures involved in actually producing music [...] and also the fictional movements and gestures of the virtual environment which they conjure up (Clarke, 2005; p 74).

In light of this, Todd and Feldman's findings make sense. As the perception of the temporal properties of sound are attuned to, or rather constrained by, real or imagined motor actions, it seems natural for the kinematics and dynamics of music, understood as perceptual realities, to obey the same physical laws as actual movements. This connection has already been proposed by Todd in the way he suggests that the phrase structure in some musical styles resembles the temporal pattern of single motor actions. The rule of "the faster the louder", and "the slower the softer" seems plausible if we consider it in connection to biomechanical constraints. In music performance, for example for a piano player, the rule holds true up to a certain point. Above a certain level of force it is difficult to perform very fast passages on a piano. This shift from a "faster to louder" to a "very fast and less forceful" rule might again be understood in terms of biomechanical constraints.

By applying the notion of music-movement relations as a perceptual reality, the question of "mass" in music also becomes easier to handle. I would suggest that the "mass" feature may be related to an imagery of how forces are applied in the movement of effectors (hands and arms, which have a certain weight) when producing musical sound on an instrument.

4.5. Dynamics vs kinematics: *activation* and *effort* re-examined

The discussion started with an understanding of dynamics and kinematics that originates from the field of physics (mechanics). At first sight it seems appropriate to apply the concepts to body movement, by understanding the movement that we see with our eyes (velocity, trajectory in space) as kinematical qualities, and the forces distributed in time (coordinated muscular activity, gravity, frictional forces) as dynamical aspects. However, when we observe the movement from outside, and consider it a perceptual experience, it appears to be more difficult to distinguish clearly between the two kinds of qualities. This is a challenge when trying to analyse body movement; what exactly in the movement that we observe can we unambiguously categorize as dynamics? And how can this be distinguished from the kinematics?

In the following I shall re-examine the concepts of *activation* and *effort* in light of the dynamics-kinematics distinction. As noted in the review of the concepts, they are both connected to dynamical aspects of experience. However, applying the understanding just outlined above that dynamical and kinematical aspects are tightly interwoven in experience, it seems that both activation and effort as dynamical features are experienced on the basis of their observed kinematical counterpart.

To illustrate this we may reconsider the example of the boy hitting his toy, and the mother joining in with a voiced 'kaaaa-bam'. Theoretically, the kinematical aspects of the boy's movement are connected to the trajectory in space and the changes of velocity. The changes in velocity may be measured and graphically displayed as a curve. However, building on a motor theory of perception, an integral part of observing these changes in velocity as they appear in the movement is to imagine how the boy's hits are produced with forces, and how these forces (muscular, gravitational) work on mass (body, arm). Apparently, this motor imagery includes dynamical aspects. Thus, it seems to me that the curve that initially represented velocity changes might as well be understood as a coarse representation, or a visualisation, of the way forces are distributed in time, i.e. the dynamics of the movement.

In chapter 3 I mentioned an experiment that demonstrated a relatively robust correspondence between 'ascending-descending' pitch contours and 'up-down' movements in space (Lipscomb & Kim, 2004). According to our definitions of dynamics and kinematics this match should be understood as kinematical (in terms of trajectory). This kind of correspondence is also demonstrated in the boy/mother example, i.e. in the way the boy's falling arm movement matches the falling pitch contour of the mother's 'kaaaa – bam'. (In chapter 6, in the analysis of sound-tracings, I shall address this kind of correspondence further.)

So, the question is whether this correspondence may be viewed from a dynamical perspective as well. Applying the understanding of motor imagery as an underlying resource of perception, to see something that moves upwards in space involves imagining the forces that are involved; in movement an ascending movement implies that muscular force needs to be increased in order to defy gravity. True enough, on some musical instruments, such as brass and woodwind instruments, applying increased pressure produces a higher pitch. However, this dynamical aspect is not relevant for many other instruments.

A further interpretation of pitch change in terms of dynamics might be found in the relation between pitch level and perceived loudness. For sine waves it has been shown that high-pitched tones (e.g. frequencies around 2000 Hz) appear louder than low-pitched tones (e.g. frequencies around 50 Hz) although they are similar in terms of sound pressure (Mathews, 2001). Psychoacoustical research of this kind suggests that the auditory system is more sensitive at middle to higher frequencies than at lower frequencies. In cases when a high-pitched tone is actually perceived as louder, this increased loudness may contribute to an increase in perceived overall intensity (see definitions of the term *intensity* below). This change in intensity may in turn, through motor imagery, be associated with increased force, i.e. a change that is linked to dynamical aspects.

This issue concerning overlaps between dynamics and kinematics is also highly relevant with respect to Laban's concept of *effort*. Initially, the effort concept appears to refer to both dynamical and kinematical qualities of the movement; e.g. the *Time* factor seems to deal with changes of velocity, i.e. kinematics, whereas the *Weight* and the *Flow* factors seem to have something to do with the way forces are distributed in time, i.e. dynamics. Although the term effort itself alludes to the dynamics, i.e. the forces that cause and constrain a movement, the motion factors as analytical and descriptive categories appear to be a mixture of both dynamical and kinematical aspects. In light of this, a strict, physical definition of *dynamics* seems to be too narrow to fully understand the way effort elements are analysed within the Laban Motion Analysis framework, and a broader definition of dynamics is required. Laban himself understood effort as the origin of any human bodily movement, the inner quality of the movement (Laban, 1971). Thus, we have to extend the understanding of "origin" from mere physicality (muscular actions and coordination) to the mental attitude that motivates the movement. For example, it might be possible to imagine two variants of 'quickness', in which one is perceived as smooth and calm and the other more nervous and hurried. In sum, this means that effort refers to a compound of "forces", physical and motivational, as well as intentional and emotional, that motivate, cause and constrain movement.

In light of this, the distinction between dynamics and kinematics appears to be more a matter of the perspective we take when observing a sonic or gestural event, rather than distinct physical categories. Underlying this view is the basic notion of motor theory that motor imagery is an integral part of both visual and auditory perception. Thus the mother's voiced response to her son's recurrent hits is based on the way she imagines that her son's movements are produced, i.e. by empathetically imagining what it would be like to perform the movements herself. This includes imagery of a compound of muscular and intentional/emotional forces, and the way these are patterned in time. This imagery of the howness of the hits is transformed into a sonic expression, the 'kaaa-bam', and in this transformation lies the possibility of an emergent audio-visual correspondence, as imitation or affective attunement. This imagined howness may also include features associated with effort elements. For example, the mother may be sensitive to whether the hits are performed with a firm *Weight*, i.e. that a sense of weightiness is fully released into the strokes, and this weightiness may also be reflected in the sound, i.e. in the "bam".

4.6. An approach to analysis based on *activation* and *effort*

The applicability of activation and effort as analytical/descriptive terms will be further discussed in chapter 7. As mentioned in the introduction to the present chapter, I have chosen them as a point of departure for music-movement analysis because I believe that an analytical approach requires a conceptual framework that addresses the way features change over time, both in a musical as well as in a gestural process. In this regard, it should be mentioned that there are alternative terms that might have been employed. For example, terms such as *tension*, *energy* and *intensity* are commonly associated with dynamical aspects of temporal phenomena. In musicological discourse *intensity* refers, first, to the psychoacoustical understanding of the relation between intensity and dynamics (dynamics understood as loudness), i.e. sound pressure measured in decibels (Rasch & Plomp, 1999). Second, *intensity* is understood as a multidimensional feature, an emergent feature resulting from many interacting features, such as dynamics (loudness), pitch range, note density, harmony (consonance vs dissonance), articulation and timbre (Krumhansl, 1996; Vines et al., 2005). As such, intensity is considered to be a prominent quality of musical experience; i.e. the temporal patterning of changes in intensity contributes to musical structure (phrasing, melodic segmentation), as well as to shadings of expressivity in music.

Furthermore, *intensity* and *tension* are often discussed in musicological discourse as features of musical performance that link music to movement. We may think of a musical phrase in which the intensity changes from a low level at the beginning towards a high level, and that the phrase is closed by returning to the low level of intensity, experienced as a point of rest. This process alludes to movement as it may be understood, at least on a metaphorical level, as a "journey" from a moment of stability, to a point of high intensity characterised as unstable, and then a return to equilibrium (Snyder, 2000).

Finally, and as noted previously, *intensity* is one of the perceptual attributes that is considered to be amodal; hence the concept might as well replace or be used interchangeably with the concept of activation. Although *energy*, *tension* and *intensity* will occasionally be used in this thesis to allude to the same features in both music and movement, I would prefer to use *activation* as the core concept. The reason for this is that activation in a very direct manner addresses the motor imagery that is thought to be an integral part of music perception and movement perception, and on which I presume audio-visual correspondences are based.

In chapter 7 I shall develop an approach to analysis that is based on *activation* as a key concept. Building on the notion that perceived 'intensity' is an emergent quality of co-evolving features, I shall apply a similar multidimensional understanding to activation as a phenomenal feature of both music and movement. With respect to music I would suggest that the following features contribute to perceived activation: density, pitch variation/change, dynamics (loudness), timbre, articulation and texture. Similarly, I would propose that density, extension, involvement, articulation and force affect the way we experience changes in activation in movement. These observational categories will be further explained and discussed in chapter 7.

Regarding the concept of effort, this will be applied to analysis primarily by implementing the understanding of movement as a multidimensional phenomenon. Building on Laban's notion of *basic efforts* (such as 'slashing', 'gliding', 'pressing'; to be explained in more depth later) and the way these are composed of co-evolving effort elements, I will attempt to develop a toolset of movement metaphors that articulates shadings of the way music and movement are characterised by changes. For example, the terms 'pushing', 'shoving', 'thrusting' and 'pressing' are all movement descriptions that refer to a similar kind of activity, but that differ slightly in the way they are performed, e.g. a ‘thrusting’ movement is executed with a quicker *Time* effort than is a ‘pressing’ movement.

The starting point of this chapter was the assumption that an analysis of music-movement relations has to address the way music and movement are understood as processes, and as such characterised by changes in features over time. I have done this by reviewing two concepts, *activation* and *effort*, that both attempt to deal with temporal phenomena in the way they focus on the howness of movement.

Another underlying assumption of the chapter has been that changes in a musical or gestural processes may on the one hand be approached and analysed in terms of dynamical features, and, on the other hand, in terms of kinematical ones. Therefore, in the second of part of the chapter I set out to clarify the way dynamics and kinematics apply to music and movement respectively. I arrived at the conclusion that, although they may easily be divided on a theoretical level, the distinction becomes blurred when dynamics and kinematics are approached on a perceptual level. This view was exemplified in a re-examination of the concepts of effort and activation.

Inevitably, discussing music-movement relationships according to the dynamics-kinematics distinction brings up the broader discussion of music and motion, or music *as* motion. First, understanding motor imagery as an integral part of both music and movement perception, i.e. that auditory and visual perception utilises the same psychological processes, leads to the notion that music relates to motion as a perceptual reality. This view was first introduced in chapter 3 and further elaborated on in this chapter. Second, in my view this notion may be further explored by examining music-movement correspondences on the basis of concepts such as activation and effort.

Exploring the way music and movement are characterised by changes brings up the issue of non-symbolic or pre-verbal meaning. Especially activation, and the way it is understood as an underlying feature of vitality affects, is embedded in a theory of non-symbolic meaning. But also effort, as a compound of physicality and intentionality, borders on issues concerning meaning in movement. The distinction between symbolic and non-symbolic meaning will be the main theme of the next chapter.

Chapter 5. Symbolic and non-symbolic meanings

> Thus, audiovisual analysis is *descriptive* analysis; it should avoid any symbolizing interpretation of a psychoanalytic, psychological, social, or political nature. Interpretation may of course follow, based on the findings of the analysis. Here, for example, it is not the symbolism of water and waves that interests us, but rather the wave as a dynamic model (Michel Chion in *Audio-Vision. Sound on Screen*; p 198).

5.1. Introduction

The point of departure of this chapter is the notion that to perceive is to understand (Noë, 2004). In the previous chapter the twin concepts of *dynamics* and *kinematics* were defined and discussed in relation to the concepts of *activation contour* and *effort*. Seemingly, the review of these concepts bordered on issues concerning the emergence of meaning. In this chapter I shall refer to features such as *dynamics*, *kinematics*, *activation*, and *effort*, as well as *chunking, peak-structure* and *goal-points* as *non-symbolic* aspects of meaning. I have previously used the terms *symbolic* and *non-symbolic* to demarcate the focus of my thesis of music and movement correspondences, and stated that correspondences will be studied primarily with respect to non-symbolic aspects of music and movement. The aim of the present chapter is to work out a perspective on the way these non-symbolic features appear as meaningful to us, despite their inability to convey referential or symbolic meaning.

This means that the distinction between symbolic and non-symbolic aspects of meaning will be the main emphasis of this chapter. To illustrate the distinction let us consider a short section from Milan Kundera's novel *Immortality*. In the introductory chapter the author describes a scene from a public bath. An old lady is taking swimming lessons; her attempts are clumsy. Her teacher is a young male student, who patiently guides her through the lesson. The author observes what is going on from a deckchair nearby. He returns to his own thoughts, but when the lesson is over and the lady climbs out of the pool and heads for the changing room, a movement of her hand draws his attention:

> She walked around the pool towards the exit. She passed the lifeguard, and after she had gone some three or four steps beyond him she turned her head, smiled and waved to him. At that instant I felt a pang in my heart! That smile and that gesture belonged to a twenty-year-old girl! Her arm rose with bewitching ease. It was as if she were playfully tossing a brightly coloured ball to her lover. That smile and that gesture had charm and elegance, while the face and the body no longer had any charm (Excerpt from Milan Kundera's novel *Immortality*,1989; p 3).

What Kundera depicts is a waving gesture, seemingly floating up into the air with a lightness and delicacy that contrasts with the clumsiness and weightiness of the woman's body. From one angle, one could suggest that the gesture bore the *emblematic* meaning of ”goodbye for now – see you next week”. But what catches the attention and inspires interpretation in Kundera’s text is not this ‘permanent’ and non-redundant meaning, it is the kind of meaning that goes beyond semantics and the physical appearance of the gesture; it is the features of the movement that leave an imaginary trace in the air, aspects that enable the observer to interpret emotions and intentions, as well as reflect on the relation between the old lady and the young man and the narrative they are taking part in.

The example illustrates how a single gesture offers a wide range of meanings, from the *symbolic* to the *non-symbolic*. The gesture’s meaningfulness ranges from meanings that are associated with words, as well as with other types of signs, to meanings that are less readily captured by words but that acquire meaning in the context of performance and participation. In light of this, I would first propose a distinction between the ‘what’ and the ‘how’ aspects of a gesture.

The chapter has been structured so that I first outline perspectives concerning symbolic meaning. This will be performed with reference to examples, such as the hand-gesture performed by the old woman in Kundera’s novel. The brief review introduces terms such as *denotation* vs *connotation* and *linguistic* vs *non-linguistic* and is followed by a more extensive discussion of non-symbolic aspects.

One of the main assumptions of this thesis is that the participants in our observational, free dance-movements study respond relatively directly to dynamical, kinematical and rhythmical qualities in the music, and that the question of music and movement correspondences should be analysed and discussed in terms of such aspects. Accordingly, it is suggested that the dancers/respondents do not assign symbolic meaning to the music via interpretation and then translate it into a similar symbolic gestural expression. This implies that the movements are considered to be pre-dominantly meaningful by virtue of their non-symbolic aspects, rather than their denotative/connotative aspects.

A similar perspective on audio-visual analysis is proposed in the epigraph of the chapter where Michel Chion favours a descriptive analysis that addresses “the wave as a dynamical model”. The aim is not to provide interpretations on a symbolic or psychological level of meaning. Nor would it seem that his aim appears to be to to account for the way socio-cultural codes and historical context are constitutive of meaning.

Discussing meaning on the basis of a distinction between opposing aspects suggests a dichotomy, and understanding phenomena in terms of dichotomies tends to rely on a simplification. The initial, separate reviews of the symbolic and the non-symbolic lead to a discussion of the way these different aspects of meaning relate to each other. In fact, I think that the analyses that illustrate Chion’s view on audio-visual analysis (e.g. the analysis of the introductory scene to Ingemar Bergman’s *Persona*[28]) show that there is a short step from his technical descriptions to a discussion of how the distribution of synch-points contributes to the emotional meaning of the scenes. This leads to the suspicion that the distinction between the wave as symbol and the wave as dynamical form (i.e. understood as a non-symbolic aspect) may not be that clear-cut, and that the analysis may suffer from an

[28] Bergman, I. (1966). Persona [DVD]: Tartan.

oversimplified dichotomy. Hence, the initial hypothesis that it is relevant to focus one's analytical attention on non-symbolic aspects may be challenged by the notion that symbolic and non-symbolic aspects should not be separated. An analytical approach that exclusively concentrates on dynamic-kinematical and rhythmical aspects may be questioned as it does not consider socio-cultural factors that play an essential, contextualising role.

In light of this, a final central issue to be discussed in this chapter is whether it is justified to study music-movement relations in the video-recorded dance-movements by focusing exclusively on non-symbolic aspects without taking into consideration that different aspects of meaning may be evoked in a single movement, as well in the corresponding musical flow.

5.2. Symbolic meanings in a single hand-gesture

As a starting point I shall return to the old woman's hand-gesture described by Kundera. The example illustrates how body movements are meaningful in many different ways. First, we may distinguish between 'what' the gesture *is* in terms of the gesture symbolising the meaning 'good-bye', and 'how' the gesture is performed.

Webster's dictionary defines a symbol as "something chosen to stand for or represent something else, usually because of a resemblance in qualities or characteristics; an object used to typify a quality, abstract idea etc".[29] In some cases, e.g. most of the words used in a written/spoken language, the symbol does not resemble the object in terms of qualities, so that the symbol is characterised by its arbitrariness. This means that the form or structure of the symbol is arbitrary vis-à-vis its meaning. For example, the word "car" as a sign is independent of the object it refers to. But whether arbitrary or non-arbitrary, a fundamental feature commonly agreed upon is that the symbol acquires its meaning by convention, as a result of historical and socio-cultural factors, and is understood by a community of users. Although it may be difficult to decide whether the form of the old lady's hand-gesture is arbitrary or whether a historical examination might reveal a plausible epistemological origin, the gesture is a sign that is recognised and assigned a relatively restricted meaning by a community of users. By convention, the symbol denotes a concept or an idea, in this case "goodbye".

Initially, I shall understand this symbolic meaning as the 'what' perspective of meaning, i.e. the meaning that is related to the typical movement trajectory that is commonly recognised as 'waving goodbye'. From the viewpoint of semiotics this restricted meaning would be referred to as the sign's denotation; the literal or core meaning of the sign, the "obvious or 'commonsense' meaning of a sign"[30]. With respect to linguistic signs, the word's denotative meaning refers to its dictionary meaning.

[29] Webster's Comprehensive Dictionary. Encyclopedic Edition (2003)
[30] David Chandler: *Semiotics for Beginners*, see
http://www.aber.ac.uk/media/Documents/S4B/sem06.html

5.2.1. Denotation vs connotation and linguistic vs non-linguistic

The sign's *denotation* is distinguished from its *connotation*, the latter term referring to meanings that are more dependent on context and personal/individual associations; i.e. the meanings that arise from a specific situation (within the course of a narrative, and the individual predispositions of the participants), as well as those meanings that are supplied by broader socio-cultural and historical contexts. The hand-gesture of the old woman is perceived to be performed idiosyncratically ("Her arm rose with bewitching ease"), and it appears in a specific context, both in terms of the narrative (e.g. the relationship between the old lady and her teacher), as well as in the bodily context of the gesture ("That smile and that gesture had charm and elegance, while the face and the body no longer had any charm"). Both the performance and the context in which it appears thus both contribute to the emergence of connotative meanings.

Having outlined the basic characteristics of a symbol and what is implied by the terms denotation and connotation, I shall now briefly present the basic distinctions regarding gesture and meaning that have been proposed by researcher of non-verbal communication. Pioneering researchers within this field, such as Adam Kendon and David McNeill (Kendon, 2004; McNeill, 1992), have worked out a theory that attempts to understand the way different kinds of gestures acquire meaning and how they become an integral part of social discourse. Starting with studies of sign languages used by deaf people, increasing attention has been drawn to the kind of free gestures that seems to accompany speech, what are known as co-verbal gestures, and which are exhibited by most people to some extent. Spontaneous, speech-accompanying gestures are here used to mean movements mainly by the hands/arms but also other body parts, and which are performed during speaking. They may be observed while people are telling a story or simply talking to each other. It is a kind of accompanying gesture of which we are not always consciously aware: we may emphasise words or parts of sentences by pointing, or we may shape the outlines of objects and imitate actions and trajectories with our hands.

It has been suggested, initially by Kendon, that the variety of gesturing may be classified along a continuum ranging from the coded gesture-forms of a sign language to free, spontaneous and highly context-sensitive speech-accompanying gestures (McNeill, 1992; p 37). In table 2 I have summarised the main ideas of this classification according to a continuum between the linguistic (lower part) and the non-linguistic (upper part)[31].

The kinds of gestures in the upper part of the table, referred to as *co-verbal gestures*, are characterised by their being non-linguistic because they produce meaning differently compared to the coded signs of a sign language. The form of the gesture is often closely related to the meaning it is intended to depict. For example, while telling someone what it was like to walk up a hill the accompanying hand movement may assume the trajectory of an upwards movement. In another example, a movement may refer to the statistical increase of a company's income. This illustrates

[31] McNeill calls the continuum from the non-linguistic to the linguistic "Kendon's continuum" since Adam Kendon first proposed this notion. My review is built on the way McNeill explains the two extremes in terms of producing meaning, see McNeill, 1992; pp 37 – 72.

that such co-verbal gestures are not part of a sign system that by convention and through learning is understood by a community of users; on the contrary, the meaning of the movement emerges from the context which is mainly given by the co-occurring speech. Thus co-verbal gestures alone do not possess distinct meanings; they are considered to be 'global-synthetic', i.e. they acquire their meaning through a synthesis determined by the global meaning of the context in which they appear.

Category	*Example/description*	*Interaction with speech*	*Non-linguistic vs. linguistic properties*
Free gestures	Making an upward motion with a hand and at the same time telling about how someone is walking up a hill	Always	• Global-synthetic • Non-combinatoric • Context-sensitive • No standards of form • Timing
Pantomime	Depicting objects or actions by mimicking gestures that are directly associated with these objects and actions.	Little	• Can be combinatoric
Emblems	Conventional signs, such as a thumbs-up gesture, a waving good-bye gesture or a toss of the head. Stand-alone gestures not included in a grammatical structure	Little	• Standards of well-formedness
Sign (as in sign language)	The sign for car in Norwegian sign language: holding the hands as though holding the steering wheel with both hands	None	• Segmentation • Compositionality • Lexicon • Syntax • Arbitrariness • Utterance-like timing • Standards of form • A community of users

Table 2. The table gives an overview of the continuum *free gestures - pantomime – emblems – signs (sign language)*. The *free gestures* are characterised by the way meaning emerges in the interplay with the context, whereas the signs of a sign language at the other end of the continuum acquire meaning by convention.

The hand-gesture of the old woman in Kundera's novel may be classified as an *emblem*. To some degree the gesture meets the requirements of a coded sign as it adheres to standards of well-formedness (the kinematical form of a gesture of waving goodbye), in the way the form of the sign appears to be arbitrary vis-à-vis what it signifies, and finally, the way it denotes a restricted meaning for a community of users, as discussed above. However, considering the way it is used in this event, it is not fully linguistic as it does not appear in a linguistic syntax in combination with other lexical signs, i.e. combining a sequence of signs into a sentence according to certain rules.

In addition to this emblematic semi-linguistic feature, I would suggest that the hand-gesture may afford further connotations by the way the gesture appears in a narrative context. Occurring as a dramaturgical element, the gesture may be assigned a meaning that specifies the relation between the old lady and the young swimming teacher; it may serve the purpose of characterising the old lady as a person, and, from the perspective of the author the gesture may have been introduced as a method of story-telling.

5.2.2. Symbolic meanings in a dance performance

How does this apply to movements in a dance performance? With reference to the ballet, *Rooster*, I would suggest that the meanings of dance-movements may be understood within the same framework. The ballet was choreographed by Christopher Bruce and first performed by *Ballet du Grand Theatre de Geneve* in 1991[32]. The performance illustrates that an analysis of meaning may be approached from many different angles depending on the kind of context we are focusing on. Seemingly, discussing the meaning of movements in a multimedia performance opens up for additional contexts with which the body movements interact and by which they are constrained.

The opening section of the ballet is choreography set to music by The Rolling Stones, the song *The Little Red Rooster*[33]. The music is characterised by its overall rhythm and blues style as well as its slow, sliding rhythmic quality. The choreography is characterised by the way a few movement patterns are varied. At the beginning, two main gestural motifs are introduced: the first motif is a movement pattern that resembles the walk of a cockerel, a pattern that a commentary refers to as a 'rooster strut'[34]: a slowly gliding, stretching, forwards-moving leg motion, the foot sliding on the floor, is followed by a quick, sudden and forceful upper-body movement. This pattern is alternated with or broken up by quick head movements as well as short, abrupt jumps.

The other main motif, preserving the quick, abrupt feature of the rooster strut, mimics the typical movements of a young man who is going out on a Saturday evening, dressed up, and on the lookout for women to conquer. The male dancers straighten their cuffs and sleeves, and they repeatedly adjust their ties. These patterns are performed quickly and hurriedly, while maintaining a vigilant face and body attitude.

With respect to the division between denotation and connotation, the motifs may on a denotative level be seen as signifying the core meanings of a strutting cockerel and a young man going out. Additionally, a wide range of connotations may be derived from the two motifs, both in terms of what they signify as symbols alone as well as how they interact with each other and with the music.

---

32 Bruce, C. (1994). Rooster [Television]: Danmarks Radio/ABCKO.
33 The Rolling Stones Singles Collection: The London Years [Box set]. Audio CD. London 2001
34 Rambert Dance Company web-site http://www.rambert.org.uk/archive/repertoire/detail.asp?art=451

The cockerel may be seen as a symbol of masculinity. The cockerel represents features that may be associated with male behaviour, i.e. the cockerel like many other male birds 'dresses up' in bright and coloured feathers to impress the females, and cockerels fight each other for the same reason. These connotative meanings may on the one hand be understood as a relative fixed symbol in our culture, the cockerel representing masculinity. On the other hand, the combination of the two main movement motifs may be seen as affording a metaphorical transfer between cockerel and man ("the man is a cockerel"), so that a variety of additional meanings may emerge (Lakoff & Johnson, 1984).

A further analysis of meaning might identify a variety of contexts, all of which contribute to different connotations of meaning, as suggested below:

- The music of The Rolling Stones, which alone calls for a range of contexts: the role of this style of music for identification processes within adolescent culture of the sixties; the slow pace of the groove, the singer as a model for identification.
- The coupling of the main motif, the 'rooster strut' with other movement motifs, such as the tie adjustments, as well as movements and postures that allude to the swing/rock'n roll dance-style introduced in the fifties. Accordingly, the dance-style also suggests the way Elvis Presley characteristically moved his hips and legs on stage, which may have connotations with the way this dance style symbolised young people's rebellion against their parents.
- The characteristic costumes, the jacket and tie, may evoke more individual contextualisations determined by personal knowledge of the historical period that the performance is referring to. In a commentary, it is proposed that the ballet reflects the choreographer Christopher Bruce's own experiences of growing up in England in the sixties[35].

In table 3 I have tried to visualise how I so far understand the relations between the terms that have been discussed. The understanding is represented as a continuum between the symbolic and non-symbolic, or with reference to Michel Chion "the wave as symbol" on the one hand and "the wave as dynamic form" on the other. In the overview I have also included the 'what' and 'how' distinction. The 'what' perspective is related to the denotation of a gesture, i.e. the initial movement pattern in *Rooster* denoting the walk of a cockerel, or the hand-gesture of the old lady denoting the meaning of saying goodbye. The 'what' aspect also has some affinity to the linguistic side of the linguistic/non-linguistic continuum, e.g. the examples mentioned above demonstrate semi-linguistic properties as they both acquire meaning from conventions.

[35] http://www.rambert.org.uk/archive/repertoire/detail.asp?art=451

<table>
<tr><td>Symbolic meaning</td><td></td><td>Non-symbolic meaning</td></tr>
<tr><td>“What”</td><td>“How”</td><td></td></tr>
<tr><td>Denotation – literal meaning, primary meaning, commonsense, conventional, lexicon</td><td>Connotation – context-dependant meanings (how, where, when, who)</td><td>Shaping, patterning, groove, effort, activation/intensity, intention, affects</td></tr>
<tr><td colspan="3">Linguistic ⟷ Non-linguistic
Sign language – pantomime – emblem – co-verbal gestures</td></tr>
<tr><td colspan="3">Indirect, mediated, cognised affective ... Direct, recognised,</td></tr>
</table>

Table 3. The overlapping ‘what’- and ‘how’ perspectives.

As indicated by the two overlapping ovals the ‘how’ aspect is located somewhere between symbolic and non-symbolic meanings, thus representing a kind of grey zone between the two main categories. Related to symbolic meanings the ‘how’ aspect, together with ‘where’, ‘when’ and ‘who’, offers a spectrum of connotative meanings. These meanings are evoked or offered by the way the individual performance of a particular gesture contextualises the gesture, e.g. the way the floating, ‘airy’ quality of the old lady’s hand-gesture appears within the narrative context and describes the relation between the lady and her swimming teacher.

To sum up, I would suggest that the symbolic level of meaning should be understood as a spectrum of denotative and connotative meanings that arises from cultural or other contextual constraints and that is readily attached to or expressed by concepts, symbols or signs (words, kinesic forms, images). In the overview I have also suggested that symbolic meaning may be characterised as indirect, mediated and cognised, and that these features distinguish the symbolic from the non-symbolic. This distinction will be further discussed in the concluding part of the chapter.

The issue of music and meaning is central in musicological discourse. Because of its transitory nature, music has been assigned an exclusive role among the arts. For example, compared to the words of a novel or the images of a film, music is not capable, at least not in the same direct manner, of distinctly representing objects and events. And this seems to contradict what is required by a symbol, i.e. to stand for something else. The less direct relation between music and symbolic meaning has imposed a major challenge on music theory; certainly, we experience music as meaningful, but as long as it is difficult to capture this meaning in words and concepts, it is hard to understand exactly how this meaning emerges.

Despite this, by applying the theory of semiotics and linguistics just outlined to meaning in music, it may be suggested that there are traces of symbolism in music. In rare instances music appears in specific context so that emblematic meanings emerge by convention, e.g. the reveille at a military camp telling the troops to get out of bed, or the national anthem symbolising the unity of a nation. In these cases the meaning is relatively distinct. However, it is clear that musical sound does not, unlike the words of a spoken language or the signs of a sign language combine according to a grammar to produce meaning in the linguistic sense. Musical meaning seems to be highly context-dependent, and it is hard to identify single elements of a musical style that directly symbolise meanings. On the contrary, the musical elements that constitute a style appear as a whole together with codes of clothing and behaviour that allude to a wide range of connotations. Accordingly, the meaning-content is characterised by its non-distinctiveness, i.e. as a spectrum of connotations not readily captured in a few words, but a set of ideas, values, concepts and symbols suggested by a musical style embedded in a musical culture.

These brief comments are not meant to initiate a broader discussion of this central issue concerning music and meaning. My point of departure for the proceeding discussion is that the rhythmical character (the 'groove') that characterises The Rolling Stones tune used in *Rooster*, on the one hand may be approached from the way it connotes symbolic meanings contextualised within a socio-musical culture, and on the other hand the way it may be understood as expressing, or performing, non-symbolic meaning.

5.3. Non-symbolic aspects of meaning: from continuous streams to coherent units

Non-symbolic aspects of meaning shall be further explored in the following. I shall return to the 'how' features; the lightness of the old woman's hand-gesture, performed with "bewitching ease", or the rooster strut characterised by its slow/gliding and quick/sudden patterns. Fine-grained nuances of movements, musical grooves and phrases affect us directly and become meaningful to us although they are not readily captured by words or concepts. They do not convey distinct meanings in the symbolic sense, but are still meaningful; they will therefore be referred to as non-symbolic.

To begin with, I shall concentrate on the old woman's hand-gesture as it appears in the course of a few seconds. Within this restricted time-window the gesture may be described in terms of a physical signal; the trajectory in space, accelerations/decelerations and velocities at given moments. With the proper equipment for tracking these aspects the gesture might be measured and represented by graphs. Selected body-parts, for example the hand, might be tracked to enable a graphical representation of the displacement in space as a function of time. Based on the tracking, a velocity curve and a corresponding acceleration/deceleration curve might be calculated. In accordance with the distinction between physical parameters and perceptual qualities, proposed by Gibson (Gibson, 1986), I would suggest that these physical parameters, measured and calculated from the signal, do not possess or represent meaning. This implies that meaning is connected to how the hand-gesture appears to the perceiver; or to use the Gibsonian term, meaning is related to the way the hand-gesture *affords* meaning to the perceiver.

5.3.1. Grouping and chunking

The discussion of how the hand-gesture is meaningful in the non-symbolic sense starts at the level of *grouping* or *chunking*, i.e. meaning emerges only when we understand the described trajectory of the hand as one single, coherent event. This means that we perceptually organise the continuous stream into more solid units; i.e. the gesture is perceived with a start and an end-point in time, the gesture is recognised as an individual event and as such distinguished from other gestures, and we have a sense of transition from one gesture to the next.

In gesture research this kind of temporal unit is referred to as a *gesture unit* (Kendon, 2004; McNeill, 1992). The hand-gesture depicted in Kundera's novel is one example of a gesture unit. Another is the gliding, forwards-moving leg motion succeeded by the more rapid, forceful body movement, referred to as the rooster strut. This single unit is repeated a number of times, and then succeeded by a new unit, e.g. the tie-adjusting pattern.

Within the field of music perception, issues concerning grouping date back to the way the Gestalt psychologists addressed the issue of melody perception as a "Gestaltqualität" (Eichert et al., 1997). In everyday life we are presented with a conglomerate of sonic elements that through perceptual organisation are linked together so that we recognise sonic events and their relation to a sound source. Likewise, a musical performance presents us with simultaneous and successive sound elements that are brought together so that they are heard as a chord, a melody, or a rhythmic configuration. Similar to the way a continuous movement is broken up and recognised in terms of gesture units, continuous sound is segmented into recognisable musical units. Snyder understands *grouping* as "the natural tendency of the human nervous system to segment acoustical information from the outside world into *units*, whose components seem related forming some kind of *wholes*" (Snyder, 2000; p 31). Bregman distinguishes between *sequential grouping*[36], i.e. "putting things together that follow one another in time" (Bregman, 1990; p 31); and *simultaneous grouping*[37], i.e. "integrating components that occur at the same time in different parts of the spectrum" (Bregman, 1990; p 31). Within the context of the present thesis, with its focus on music and movement correspondences over time, the term *grouping* will refer primarily to sequential organisation.

Instead of grouping I would prefer in the following the closely related term *chunking*. Similar to grouping, chunking addresses the way we link elements so that they appear as larger units; however, chunking refers more specifically to the strategies we apply when we attempt to memorise sequences. For example, a succession of digits is linked together in chunks of digits so that they can be remembered as a telephone number or a datum.

[36] Also referred to as *sequential integration*, *sequential organisation*, or *stream integration* (Bregman, 1990).
[37] Also referred to as *spectral organisation* (Bregman, 1990).

> Chunking is the consolidation of small groups of associated memory elements. Just as a single number is a unit that is distinguished from all other by its unique visual features, the individual items in a chunk become its features. It is as though chunking helps memories "coagulate" or solidify (Snyder, 2000; p 54).

The key word to characterise chunking is understanding or meaning. We aid our memory by imposing a meaning, or an understanding, on the otherwise meaningless succession of digits. In other words, we fix the continuous stream into a more solid unit by relating it to some kind of pre-understanding, which means that chunking involves an element of recognition. This way of explaining chunking is in accordance with the view that perception includes understanding.

I shall use the term to refer to both music and movement chunks. To illustrate this, the series of stills (figure 10) shows a video-recorded movement from our observational study; the dancer raises her arms so that her hands meet in a position above her head. The movement appears as one unit, or one movement chunk.

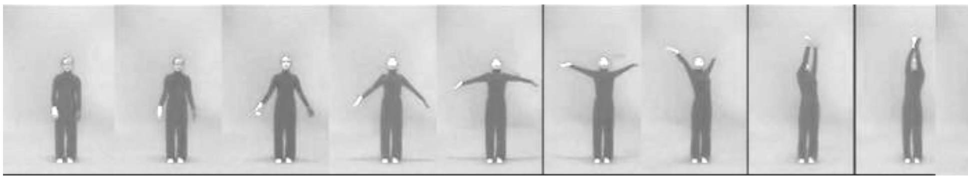


Figure 10. Raising the arms to a position above the head is seen as one coherent movement chunk, the series of stills used as a representation of the continuous movement.

This movement corresponds to a sustained timbre in the music, which is similarly heard as a musical chunk:

Figure 11. The sustained timbre, marked with a box, (the beginning of the *Lento* from Ligeti's *Ten Pieces for Wind Quintet*) is heard as one coherent musical chunk. (The score is notated in C.)

On what basis do we see and hear, respectively, these two continuous streams as coherent chunks? What kind of understanding do we apply in order to chunk the movement and the music in this case?

As a general answer to these questions I would suggest that chunking processes are based on a pre-understanding of body movement. When observing a movement, such as the one illustrated in figure 10, we know with reference to our own body how movement units tend to be initiated, how they evolve in time, and how they tend to reach *closure* (see below). In accordance with the view that music and movement are linked as a perceptual reality (cf. chapter 4), we may assume that chunking in music is constrained in the way music evokes motor imagery. By relating musical chunks to action units in this manner it is suggested that also chunking in music, as with the chunking of body movement, draws on a pre-understanding of movement processes.

The term closure refers to our perception of how a musical chunk is "coming to an end", so that a boundary between two phrases is established (Snyder, 2000). The boundary may be perceived as clear and distinct, or in other cases as more blurred. This means that we have degrees of closure. In music, closure may be established as a result of dissimilarities between two phrases, e.g. when there is a shift from one instrument to another so that the timbre and/or pitch changes. Additionally, a number of other features contribute to our sense of closure (Snyder, 2000; pp 62 – 66):

- Decrease in intensity
- Falling pitch
- Decrease in overall tempo and increasingly longer durations
- Repetition (i.e. familiarity with a phrase creates expectation of points of closure in proceeding phrases)

5.3.2. Dynamical and kinematical shapes

To further elaborate on this, I would propose that the process of chunking is closely connected to the way we perceptually *shape* continuous streams. The movement in figure 10 is characterised by a trajectory, an excursion of the two arms from a low position to a position above the head. This will be referred to as a *kinematical shape*. Since the movement appears as a slow, gliding movement, we may imagine that the movement is performed with a continuously applied force. This will be referred to as a *dynamical shape*.

Similarly, the musical chunk may emerge as a coherent unit through the way it is connected to shapes heard in the sound. For example, in this case the timbre is becoming brighter and tenser, i.e. a feature that may be related to a dynamical shape; whereas the non-changing pitch alludes to a kinematical shape.

Godøy has discussed the role of shapes in music perception. He suggests that musical objects, "understood as delimited segments of actual sonic unfolding" (Godøy, 1997) may be understood and explored in the way they evoke images of shapes, i.e. in the way "emergent qualities such as timbre, texture and contour" are characterised by the way they change over time, and by the way this change may be imagined as a shape: "[…] based on the conviction that thinking of these perceived emergent qualities as shapes is a privileged mode of representation in our minds by producing *images of dynamic unfolding*, providing more or less stable images of an otherwise ephemeral sonorous flux" (Godøy, 1997; p 89).

I shall use the term shape to refer to the perceived dynamical or kinematical form of an event, i.e. the immediate perceptual impression and recognition of a trajectory in space (kinematical shape) and the perception of the distribution of activation or force within the time-window of the event (dynamical shape). The perceptual process of shaping is immediate in the sense that we are able to intuitively mimic the trajectory or the activation contour. Furthermore, the term *shaping*, suggests that this is an active explorative process of the mind, meaning that the mind actively searches for and shapes features in terms of change. Thus, the way an event affords shaping is essential for recognising temporal unfolding as coherent chunks.

To further exemplify the notion of shaping with respect to movement, and given the description of the waving hand-gesture described by Kundera, I can imagine that the hand 'shoots' upwards and then returns to a position close to the starting point (in space). This image of a kinematical shape, the trajectory of the hand, is suggested in the drawing in figure 12:

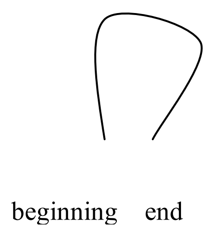


Figure 12. The drawing indicates how I imagine the trajectory, the kinematical shape, of the waving hand-gesture, as described by Kundera.

By imagining the forces involved in the performance of the gesture, I would suggest a dynamical shape something like this:

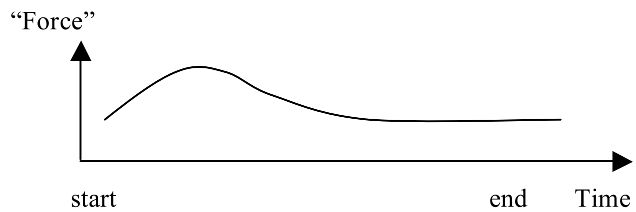


Figure 13. The curve suggests a dynamical shape, i.e. the way the waving hand-gesture may have been produced by forces (y-axis) as these are applied over time (x-axis).

The curve indicates that the movement is increasingly forceful at the beginning, that there is an emphasised peak, and that the hand floats more freely, i.e. with relatively little controlling force, towards the end. We have two shapes that describe the gesture, the first one related to the kinematics of the movement, the other alluding to the dynamics. It should again be emphasised that the shapes do not represent physical "measured" parameters. They are meant to indicate the way I, or any observer of the event, by imagining it on the basis of Kundera's words, could relate the perception of this specific gesture to two interrelated shapes. Or put differently, if I were given the task to imitate the gesture I would "think" of it as a movement that goes 'up-down' (and slightly to the side), and a movement that is most forceful at the beginning and more freely floating towards the end.

The perceptual feature of a shape is related to the way gesture researchers understand gestures as *excursions*, as well as in terms of *peak structure*:

- The term excursion refers to how a movement is leaving and returning to a rest position. Moments of rest demarcates temporally a gesture unit (Kendon, 2004; McNeill, 1992).
- It seems that people tend to perceive gestures as having a basic peak structure, i.e. the gesture shape is centred around a moment of emphasis that is preceded and succeeded by phases of no accentuation, and that this structure further shapes and solidifies the gesture as a coherent unit. In research literature on gestures the temporal patterning is prototypically referred to as the phase structure 'preparation' – 'stroke' (emphasis/peak) – 'retraction'/'recovery' (Kendon, 2004; McNeill, 1992).

These aspects of movement contribute to the process of chunking: on the one hand we have the gesture characterised as an excursion, which may be viewed as a kinematical shape; whereas the peak structure alludes to the dynamical shape of the gesture. This understanding of the way movement is temporally demarcated has been worked out by researchers of co-verbal gestures, and it should be kept in mind that the video-recorded dance-movements from our observational study may not always be described with respect to an excursion or a peak structure. For example, the movement described by the stills in figure 9 neither returns to a position of rest, nor does it evolve according to a peak structure; and yet, it can in my view clearly be perceived as one coherent movement chunk. Nevertheless, I believe the understanding and descriptive terms proposed by this kind of movement research will provide a useful basis for analysing chunking in movement.

Returning to music perception, we may think of a number of changes, such as changes in timbre, pitch, tempo, and loudness, that afford shaping, thus contributing to chunking. A succession of tones is shaped according to its melodic contour, as in figure 14:

Figure 14. The melodic contour suggests a kinematical shape that contributes to chunking.

Similarly, a crescendo followed by decrescendo affords a dynamical shape:

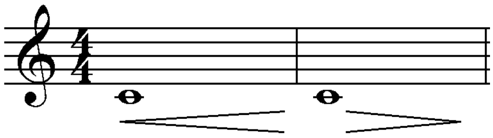


Figure 15. The *crescendo-decrescendo* suggests a dynamical shape that contributes to chunking.

In certain musical styles a typical phrase structure is characterised as a three-phase process: *pushing forward (crescendo) – climax/goal-point – release of tension (decrescendo).* I would suggest that this phrase structure in music may be understood as analogous to the prototypical 'preparation-stroke-retraction' structure in gesture. This link may be grounded in the motor theory explained in chapter 3, i.e. the perception of sonic and gestural sequences is rooted in an imagery of action units.

To illustrate this, we may imagine a musician performing the melodic contour in the example in figure 16 with this characteristic phrase structure:

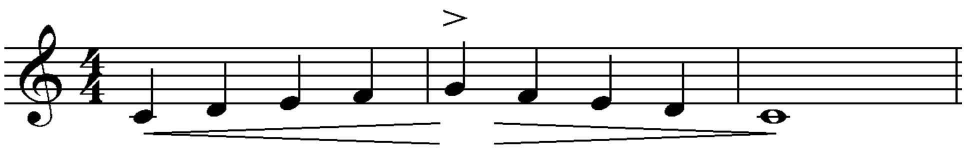


Figure 16. The kinematical shape, the dynamical shape and the shape suggested by the phrase structure work together to establish chunking.

This means that the succession of tones is characterised by two simultaneously evolving shapes, one kinematical and one that may be associated with a dynamical aspect. We may even imagine the segment played with variations in tempo to further emphasise the goal-point, so that we then have an additional shape in terms of tempo change. In this manner, changes in a number of features may interact and contribute to chunking.

The expressions *peak structure* as used in gesture studies and *phrase structure* (alternatively *phrasing*) in music needs to be further elucidated. Seemingly, the way the expressions are explained in terms of preparation/pushing forward – stroke/goal-point – retraction/release of tension suggests that an element of intention is involved. This implies that experiencing a certain phrasing, in addition to being afforded by variations in dynamics and accentuation, is affected by a 'sense' of 'going somewhere', 'reaching a goal', and 'retreating'. This intentional aspect of peak structures or phrasing is captured in the notion of gesture as an excursion. Although not easily identified in terms of specific, observable features in a movement or musical performance, I would propose that this aspect, which may be referred to as an *intentional shape*, also contributes to chunking. At this point it might be argued that with the introduction of intentions, and the way intentions are connected to metaphors, we are close to entering a more symbolic sphere of meaning. This illustrates the point that will be elaborated on later in this chapter, i.e. that there seems to be a blurred boundary between non-symbolic and symbolic aspects of meaning.

5.3.3. The prominence of goal-points

In the following I would prefer to use the same vocabulary for both music and gesture to describe such dynamical shapes, i.e. *prefix – goal-point – suffix* (Godøy, 2008b). In the discussions of audio-visual instances later in this thesis, an analysis of chunking will be addressed with respect to peak structures. From my own experience with analytical work of this kind, e.g. attempting to describe the chunks of a movement sequence, it seems that the goal-point of a chunk plays a prominent role and is easier to identify compared to the temporal boundaries between two succeeding chunks. It is as though the goal-point, temporally located somewhere in what one might call the middle of the segment, sticks out and serves the function of demarcating and solidifying a gestural or musical chunk, rather than the beginning and the end of the movement. The point is illustrated in figure 17 (see Godøy, 2008 for a discussion):

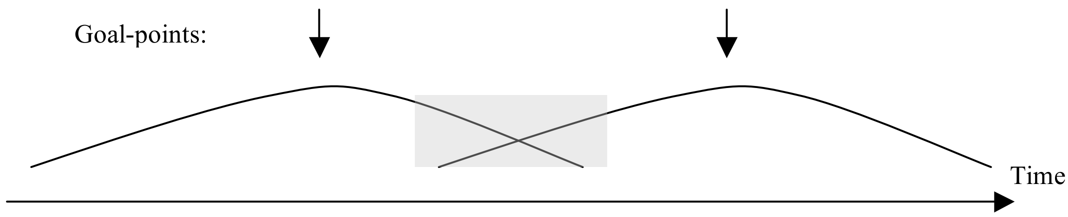


Figure 17. Schematic visualisation of two successive chunks characterised by a peak structure. The two goal-points stick out, whereas the boundary between the two chunks is less clearly defined.

This intuition seems to be in accordance with recent studies of imitated movements. In the article I briefly mentioned in the previous chapter, Wohlschläger and colleagues discuss theories that assume that imitation is characterised by a direct visual-to-motor mapping (Wohlschläger et al., 2003). Based on observations of the imitative behaviour of young children, they propose the alternative view that imitation is basically goal-directed (referred to as GOADI, i.e. goal-directed behaviour). Children who imitated a model movement showed a higher level of precision around target moments, and more approximate correspondence between these accentuated moments, i.e. what I refer to above as the prefix and suffix phases. For example, when encouraged to imitate the action of raising one's hand and touching one's ear, the moment of touching one's ear is precisely performed according to the model, whereas the trajectory leading up to the ear is much less accurately imitated. A prerequisite for this strategy is that the child is able to predict or interpret the intentional goal of the movement.

> In other words: in imitation it is primarily the goal of an act that is imitated; how that goal is achieved is only of secondary interest. Of course, perceiving the goal of an action would be a pre-requisite for such a GOADI. Indeed, recent research showed that six-month-old infants already selectively encode the goal object of an observed reaching movement (Wohlschläger et al. 2003; p 502).

Although this view is based on observations of children's behaviour, I would propose that a similar strategy may be applied in the perceptual chunking of continuous musical and gestural streams, i.e. that goal-points gain a certain prominence. However, in contrast to a relatively simple instrumental movement, such as the action of raising the hand and touching the ear, the goal-points may emerge less clearly defined in, for example, a musical sequence. Accordingly, it may be difficult to identify the specific features of a movement or a musical segment that make a certain moment appear as a goal-point, but since identifying goal-points in music and movement will be a recurrent theme in the analytical work of this thesis I will summarise what has just been said and attempt to further clarify what I mean by a goal-point:

- Goal-point is used to mean a moment within the time-window of a movement or a musical chunk that appears to be more weighted, or more distinctly accentuated/articulated, so that it contrasts with and sticks out of its temporal surroundings. Godøy proposes that goal-points may be understood as "goal postures at specific points in time", where the expression goal posture applies equally well to music when we understand the imagery of musical chunks as imagined sound-producing action (e.g. the imagery of the way the hand strikes a chord on the piano) (Godøy, 2008; p 9).
- This suggests that the perception of a musical goal-point is linked to imagery of sound-producing actions.
- In music, a goal-point will be in the form of an accent, e.g. a downbeat, or other kinds of salient moments in a melodic, dynamical or a timbral shape (Godøy, 2008).
- In a movement, a goal-point will appear as the goal posture of a movement trajectory (e.g. the action of touching one's ear in the example above). The feature of accentuation in this moment may be caused by a directional change and/or increased intensity/force.
- In music, a goal-point in many cases results from the way the musical sound emerges with a higher degree of intensity at this moment, often preceded by a build-up of tension and followed by a release of tension.
- In both music and movement, the expression goal-point suggests metaphorically the point in time that the process is 'heading towards', as well as 'pulling back from'
- We may also assume, from the perspective of the perceiver, that when we are attending to a continuous stream of music or movement we watch out for goal-points as a kind of marker or reference point on the basis of which the stream is broken up and chunked. Thus, on the basis of the observations made by Wohlschläger and colleagues we may speculate that a goal-point is a moment that receives increased attention from the perceiver.
- Another way of articulating this last point is to say that the goal-point is *afforded*, meaning that it results from both the increased physical intensity of a moment and the intention of the performer to 'reach' this moment in time, as well as the intention of the perceiver to 'screen' the continuous stream for moments that may serve as a goal-point, hence deserving increased attention.

5.3.4. Local and global features contributing to chunking

In his discussion of sequential grouping, Bregman has on the basis of empirical studies demonstrated that the perception of melody emerges from a certain interplay between pitch relations and temporal conditions, e.g. a slow tempo allows for leaps in the melody whereas a high tempo requires stepwise pitch relations to keep the tones together as a melody (Bregman, 1990). This understanding of grouping emphasises local features, i.e. the way one single tone in a sequence relates to the tone preceding it and the next one in the sequence, as well as the way a global melodic shape emerges from such pitch relations.

The role of global features has been further studied in terms of melody perception, i.e. the way the overall shape of the melody contour affects how a succession of tones is recognised as a melody (Dowling, 1994). With reference to experiments, Dowling suggests that a melodic contour appears to be a salient feature according to which respondents judge similarity between short melodies, but that melodic-contour properties interact with other features, such as tonality and rhythmic features. From the field of psycho-linguistics a similar view is proposed, namely that the prosody of verbal utterances seems to provide an overarching structure that promotes recognition, the ability to hold the auditory linguistic sequence in memory and finally the comprehension of the utterance (Frazier, Carlson, & Clifton Jr, 2006).

In my view, acknowledging shaping as a perceptual process that captures the holistic properties of a succession of tones suggests that perceiving the melodic contour or phrase structure contributes considerably to chunking in addition to the pitch relations that Bregman focuses on, i.e. that melodic and dynamic shapes are understood as immediately grasped features that in a top-down manner constrain our perception of chunks. Moreover, this implies that shapes, afforded by physical features of the perceived phenomena, from the viewpoint of ecological theory are viewed as especially useful for perceptual exploration, meaning that our perceptual system is biased towards detecting features of the shape, or the peak structure, of events.

5.3.5. Shaping categories

One further aspect of the shaping of a temporal event should be mentioned. Theories of categorical perception would suggest that the shaping of the event is contextualised by previous knowledge of such events. According to this understanding, experience is constrained by a categorisation of prototypes; we may assume that our minds distinguish between sustained events, i.e. events in which force is applied evenly and continuously, and events that are characterised by a clear emphasis. And we may assume that this last category is further divided with respect to the temporal location of the emphasis. For example, returning to the old lady’s hand-gesture, I have in the drawing of the dynamical shape indicated that the emphasis is located at the beginning of the shape, i.e. a *ballistic shape*. In the drawing next page the dashed curves suggest two additional prototypes, an arc-shape characterised by the emphasis in the middle of the segment, and a more stretched shape with the emphasis appearing towards the end of the event.

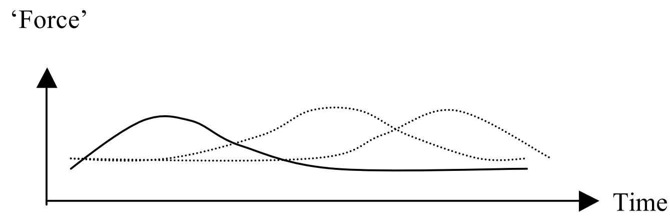


Figure 18. The solid curve indicates how forces are applied over time in a typical *ballistic shape* where the movement is produced by an initial impulse. The two dashed curves show two other categories of shapes. In some cases the experience may shift between the shape-categories, depending on the context.

Applying the understanding that is proposed by Stevan Harnad in his key concept of "approximationism" (Harnad, 1987), the shaping process of assigning the movement to one of the three alternatives is a matter of interpretation. The boundaries between the ballistic shape and the arc-shape, as well as the arc-shape and the stretched shape, may in some cases not be distinct. In such situations Harnad proposes that categorisation is context-driven, so that our experience of shape may "flip" from the arc-shape to perceiving it either as a 'ballistic' or as a 'stretched' shape. The context that constrains the perception of shapes refers both to an historical context, i.e. the way previous knowledge of events is understood by the mind as approximate patterns, as well as the context provided by the present, e.g. co-occurring events, the temporal context of the event, etc.

> The idea is that all of our categories turn out to be approximate rather than "exact" ones (in some realist's sense of "exact"); we converge on these approximations by accumulating data from experience and continually updating the categories in accordance with the constraints and contingencies of experience so as to yield an approximate match that is adequate for the sample of categorization problems we have faced to date (with the past always subsumed as a special case) (Harnad, 1987; pp 538-539).

This view may be seen as converging with one of the basic notions of ecological theory, i.e. that perception is explorative and is driven by the intention to make sense of events. The role of categorization to serve a need, i.e. to detect and understand the event, is in Harnad's terminology referred to as "optimization" or "satisficing". Again, the audio-visual example involving the steady tone that is heard with an increasing brightness when combined with the visual display of a growing shape may be re-interpreted within this framework of understanding. The context provided by the visual shape suggests that to assign the sonic event to the category of a rising dynamical contour, rather than to the category of a dynamical prototype characterised as sustained, optimizes the understanding of the fused audio-visual event.

It should be noted that in other cases the threshold between shape prototypes is more robust. One example of this is the distinction between plucked and bowed sound-producing movements on string instruments. Listeners seem to be able to identify clearly and unambiguously the point at which there is a change from one category to the other.

5.3.6. Pattern

With other kinds of musical elements the term *pattern* may be more appropriate in terms of descriptive value than shape. A succession of guitar chords is heard as a riff, and an overall rhythmical pattern with a characteristic flow quality is experienced as a groove. The riff and the groove are both characterised by repetition and contrasts of accentuation. I shall use the term pattern to refer to the way a succession of contrasts is distributed in time so that accentuated moments and moments of less intensity alternate recurrently and similarly, and so that a sense of periodicity emerges. Certainly, shape might have been used for such processes as well, but I think that pattern more directly emphasises the repetition feature.

In the video-recordings of dance-movements collected during our observational study we can also see certain movement patterns emerging. Here, pattern means that a certain movement chunk is repeated (see figure 12, where the dancer lifts her right arm out from the body and back again twice), or that a chunk alternates with another 'different' chunk in a certain manner.

Figure 19. The dancer lifts her arm slowly out from the body and back again. This movement excursion is seen as one coherent chunk and is repeated, thus inducing a sense of *pattern*.

This means that patterning is based on perceiving chunks with respect to similarities and contrasts. Similar to shaping, it is reasonable to believe that patterning as a global feature contributes to chunking in the way a perceiver 'screens' the continuous stream for similarities and contrasts, and 'evaluates' whether these similarities and contrasts may be assigned to some kind of pattern.

In sum, I would understand chunking, shaping and patterning as inter-related processes of the mind, i.e. something that the mind does to continuous streams of movement and music in order to make sense of events. This implies a pre-understanding of the features that constitute events. The notions of chunking, shaping and patterning represent a first step towards understanding how music and movement become meaningful in terms of non-symbolic aspects.

At this point it might be argued that chunking as a basic perceptual process should be understood as nothing more than a prerequisite for higher-level interpretations in terms of symbolic meanings, so that chunking should not be included in a discussion of meaning at all. However, the view that I have advocated here is based on the idea that chunking requires the perceiver actively explore the continuous stream of movement or musical sound with respect to shapes, goal-points and patterns. For this reason, since a kind of understanding is included in the perceptual process to solidify segments of continuous streams, I would propose that it is relevant to discuss chunking as an aspect of meaning.

5.3.7. Non-symbolic aspects of meaning on the basis of sensitivity and responsiveness to transformations

The next step towards an understanding of non-symbolic meanings concerns the way these dynamical and kinematical shapes become meaningful by virtue of our sensitivity and responsiveness to transformations of such shapes. To start with, I shall illustrate with an example: in a sequence from Edgar Varese's *Amerique* a dynamical theme is introduced which may be described as successive outbursts of energy, or as waves of different shapes. They are characterised by a recurrent dynamical pattern of *crescendo – marcato - decrescendo*. The crescendo and the decrescendo are more or less gradually or abruptly applied, and the range of the dynamics is also varied. First we have three waves that are played similarly, although the initial crescendo is more stretched, compared to the crescendos in the two following waves. The waveform display in figure 20 is an illustration of this first pattern[38]:

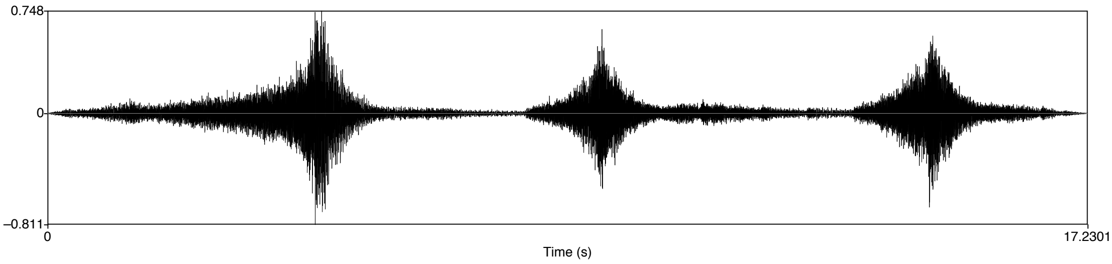


Figure 20. A waveform display of a sequence of three successive waves from Varese's *Amerique*.

Just after this a similar sequence follows, which may be seen as an elaboration of the dynamical form introduced in the first sequence. This sequence may be described as three sudden, explosive outbursts followed by a longer wave when the full release of energy is suspended, i.e. the crescendo is more stretched and a 'fake' marcato appears as a kind of violation of our expectations, as illustrated in figure 21:

[38] The display, showing amplitude variations over time, has been made using Praat software for acoustical analysis, see http://www.fon.hum.uva.nl/praat/ The analysis has been made on the basis of an excerpt from a recording of *Amerique* performed by the Royal Concertgebouw Orchestra, conducted by Riccardo Chailly (Varèse, E. (1994). [CD]: Decca.)

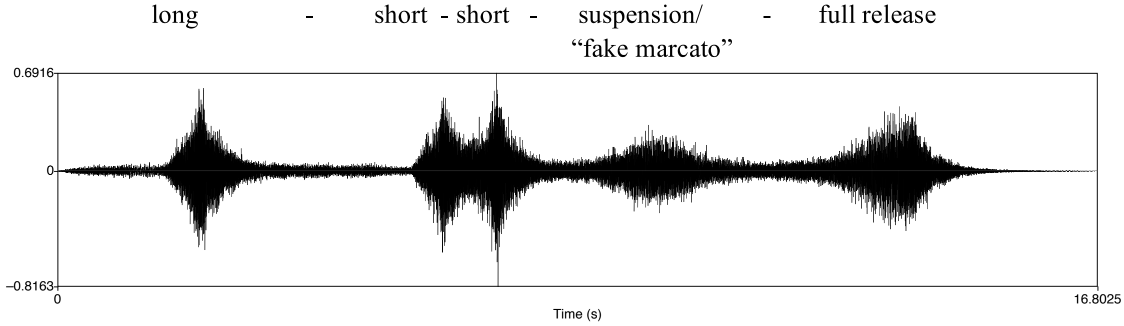


Figure 21. A waveform display of a sequence from Varese's *Amerique*. Elaboration of wave motif.

In this example, different shapes are interacting. As described above each chunk, or wave, is characterised by a crescendo-decrescendo dynamical shape; additionally, an upward-downward shape in pitch as well as a shape related to the brightness of the timbre (increase-decrease) may be identified. In this case the shape related to dynamics appears to be the most salient feature. There is also a feature related to temporal patterning, introduced in the first segment with the more stretched first wave followed by two shorter waves constituting a long-short-short pattern. This temporal patterning becomes increasingly important in the second segment when two short successive waves are compressed within a relatively short time-window. These are followed by two stretched waves that fill a much longer time span, this variant of temporal patterning at the same time interacting with variations in dynamics and articulation.

I would suggest that these waves as dynamical shapes, the variations in dynamical range (in terms of loudness) as well as the temporal distribution leave a impression that is quite essential to human experience. The experience may be likened to what we sense in our body when sitting in a small boat. We can feel the steady, rolling movements of the boat, our body responding to the waves in a similar rolling manner. Then, a change in the dynamics of the waves may occur, the steadiness is interrupted by more abrupt changes, and we adjust our own body movements to regain balance.

The wave-motif in the example from Varese may be seen as bearing meaning on a structural level; the motif is introduced and then elaborated on and varied, thus promoting musical coherence. The main point to be illustrated with the example is as follows: the waves in the musical example and the rolling movements of a small boat are meaningful to us by virtue of our sensitivity to changes and by the way we *respond* to these variations, whether the differences are subtle or more obvious. They are meaningful in the way they makes us respond and the way they evoke a sense of intention, such as pushing forward or holding back; building up energy and releasing tension.

5.3.8. Groove in music as a non-symbolic aspect of meaning

The discussion of wave shapes in the examples from *Amerique* is related to phrasing in music. As previously discussed in connection with the temporal shaping of body movements, the phrasing in some styles of music prototypically adheres to a *prefix - goal-point- suffix* structure. Another phenomenon that exemplifies non-symbolic meanings is the experience of *groove* in music.

The term groove is commonly associated with musical styles in the popular music tradition, such as rock 'n' roll, blues, funk as well as the more contemporary genres of hip-hop, modern dance etc. The term has its counterparts in other styles as well, e.g. *swing* in jazz music. These related terms refer to a certain flowing quality in the rhythm, which, typical of these musical styles, plays a prominent role; as an aesthetic quality, as a standard for the evaluation of good and bad, for demarcating the music in relation to other styles, and for the way people respond in terms of movement (dancing).

It is assumed that the groove quality of a particular performance is an emergent quality resulting from a number of co-occurring and interacting components, such as timing in relation to meter, dynamic contrasts, articulation, timbre, as well as the interplay between different textural layers or roles (e.g. melody line vs rhythmic accompaniment) (for a detailed discussion of funk grooves see (Danielsen, 2006)).

Empirical studies of rhythm have paid particular attention to the way timing properties affect the rhythmic flow. For example, the Swedish researchers Bengtsson and Gabrielsson examined the way variations in the durations of each beat within a meter resulted in different rhythmic patterns. They found, employing an analysis-by-synthesis approach that a particular rhythmic feel emerged from systematic deviations from a metrical grid. These deviations were found both on the level of a single measure, as well as on a phrase level (Bengtsson & Gabrielsson, 1983).

Bengtsson and Gabrielsson performed their analyses of performances of Swedish folk-tunes as well as on variants of waltz performances in the classical Viennese tradition. The typical timing deviation patterns were represented by graphs so that the differences of each idiosyncratic pattern were visualised and documented. Furthermore, they suggested that a specific timing pattern together with articulation features (legato, portato, staccato) affected the experience of what they referred to as the "motion character" of the rhythm (Bengtsson & Gabrielsson, 1983; p 30), an expression that may be understood as being analogous to groove or swing. They acknowledged that, although their graphs visualised differences in rhythmic patterns, they were not able to describe or capture this motion character, which is so important for the musical experience. Hence, they ended up concluding that the analysis of rhythmical phenomena in music requires new approaches in order to facilitate representations of this flowing quality (Gabrielsson & Bengtssson, 1983; p 58).

The work of Bengtsson & Gabrielsson may be seen as a follow-up to performance research conducted by Seashore and his colleagues in the 1930s (for a review, see (Gabrielsson, 1986). Examples of more recent contributions are found in studies of so-called "expressive timing" (Repp, 1998, 1999), as well as in a study that uses frequency modulation to model the timing properties of rhythmical flow (Waadeland, 2000).

Within the present discussion of non-symbolic meaning, the notion of rhythmical features and their relation to movement features is particularly interesting. As noted in chapter 1 the rhythm concept may be defined by two interrelated aspects: first, rhythm is understood in terms of a temporal ordering of contrasts; second, rhythm is understood as a certain flowing quality (Waadeland, 2000), referred to above as groove, swing or motion character. Within the corpus of rhythm research in music, it seems to be a general finding that methods for describing and analysing the first aspect, pattern properties, is relatively easily worked out, such as in graphs of deviation patterns. On the other hand, as observed by Bengtsson/Gabrielsson, these graphs do not capture the experiential qualities of motion character or groove. So, although clearly interrelated aspects, the analysis of groove is not readily accessible through mere descriptions or symbolic representations of the patterning properties.

Employing the understanding that the experience of groove is closely related to the motor imagery that the music affords the listener, or the movement pattern that is performed by a dancer, I would suggest that an analysis of grooves should take such movement qualities as its point of departure. This implies that the motion character is only to a limited extent found for example in the step pattern of the dancing, such as in different variants of waltz. The motion character is found in dynamical subtleties that constitute different styles of dance music, the way the interplay of numerous components affords different dynamical shaping patterns. For example, a waltz pattern is typically performed with a heavy, weighted step corresponding to the first beat of the measure in the music, and a lighter, uplifted step corresponding to the following second and third beats of the measure. From this basic pattern, different styles of waltz will be characterised by subtle differences in the weightiness of the first step and the quality of lightness in the second step, as well as an infinite number of variations due to how these two steps are timed and weighted in relation to each other. And these subtle variations, in addition to many other characteristics, distinguish different grooves, or motion characters, from each other.

In sum, groove as a flowing quality that is prominent in meter-based music and the nuances of wave phrases in the example from *Amerique* are here understood as a mode of experience that is highly important in the experiencing of both music and movement, and that is closely linked to bodily experience. Such musical phenomena are meaningful in terms of their motion character, the meaning attached to our direct responsiveness to changes in features, which again is based on our sensitivity to fine-grained nuances.

Although this motion character is not readily captured by concepts or other symbolic representations, it may be accessed and explored by using metaphors grounded in body experience, such as swirling, swaying, leaping, gliding etc. In this regard, the Laban Motion Analysis outlined in chapter 4 may be seen as an attempt to capture the rhythmical flow with a multidimensional approach. Returning to the old lady's hand-gesture, what catches Kundera's attention seems to be the contrast between two ways of moving, i.e. two different rhythmical flows. During the swimming lesson Kundera provides the reader with the following clues to imagining the movements: "He [the lifeguard] was giving her orders: she was to hold on to the edge of the pool and breathe deeply in and out. She proceeded to do this earnestly, seriously, and it was as if an old steam engine was wheezing from the depths of the water". Based on this, the movement may be described in the terminology of the theory of *effort* according to Laban; I imagine the moments as being controlled (*Flow*) with a direct focus (*Space*), they are slow (*Time*) and heavy (*Weight*). Contrasted to this is the waving gesture described in the introduction of the chapter; free *Flow*, indirect *Space*, quick *Time* and light *Weight*.

5.3.9. Daniel Stern: invariant patterns of awareness

Our sensitivity to such contrasts and transformations is a central issue of the ideas about preverbal meaning proposed by Daniel Stern. The reader may recall that Stern introduced the activation contour as an underlying feature of vitality affects, and that the interplay between baby and carer was based on chunks, i.e. "utterances" of relatively short duration, which are characterised by a certain rhythm or patterning of unfolding energy, and which both parties in the communication recognise and respond to. The activation contour refers to the dynamic profile of a chunk; for example, as articulated in a slow upwards movement with a hand performed by the baby, and typically, responded to with a similar, matching vocalised, upwards-pitch contour by the caregiver. This implies that the concept of activation contour corresponds to the notion of dynamical-kinematical shape*s*.

When one of these individual patterns characterised in terms of a rhythmical patterning or an activation contour is repeated in the interaction between infant and caregiver, and then introduced again the day after and a few days after that again, it eventually becomes a recognisable pattern, what Stern denotes to be an *invariant pattern of awareness*. Understood as a form of organisation, such invariant patterns may be viewed within the framework of theories of categorical perception.

The expression *invariant pattern of awareness* takes a very prominent role in his understanding of the way the infant's sense of a self is constituted. He argues that these invariant patterns become the building blocks in the formation of a preverbal self; they serve the purpose of organising subjective experience as they make something that happened to me, or to my body, in the past recognisable to what is happening to me, or to my body, in this present moment.

In the context of the infant-carer relation we may assume that such invariant patterns build a solid ground on which new variants may be explored so that the repertoire of the communication is extended. This kind of togetherness may be characterised as a series of patterns that are repeated, transformed by transfer to another sensory modality, or by slightly varying features of rhythm and activation.

All this activity is going on without the constraints or intervention of naming or conceptualisation. As is clear from this discussion these utterances are not signs in the linguistic sense. They are highly context-dependent, and their meaningfulness is not related to denotations and connotations. Hence, they do not pretend to signify; their meaningfulness is rooted in the purpose it serves for the infant to relate to its own self, and rooted in establishing and maintaining the bond and the interplay between infant and caregiver. So the meaning is implied in the way the patterns and their variants appropriate the sharing of subjective experience, and the mutual experience of knowing that we are sharing. This has led Stern to propose that what is going on in this interplay may be more precisely captured by the term *communion* than the more common *communication*, as the latter term is associated with a process of transferring a message from a sender to a receiver.

Stern's points of view on early communication and the way the interplay is based on a sensitivity to activation contours (or shapes) illustrate perspectives that appear to be relevant for a more general understanding of meaning processes in music as well as in movement, and independent of age. As Stern's theory addresses the interpersonal world of small children, the expression *pre-verbal meaning* seems to be appropriate. However, for the present I shall continue to use the expression *non-symbolic*, according to the view that such aspects of meaning also exist for those of us who have the ability to verbalise meaning.

First, I have linked the non-symbolic mode of experience to shaping, patterning and chunking processes. Second, the 'content' of non-symbolic meaning seems to be our sensitivity to similarities and differences in dynamical-kinematical shapes, as well as in the way these shapes are transformed and varied in temporal sequences. This sensitivity is articulated and further elaborated in the way we participate in such events, i.e. in our responses to similarities, differences and transformations. To summarise, the wave as a dynamical form, the waving hand-gesture of the old lady, and the groove in a certain piece of music, all become meaningful in the following manners:

- We *respond* to the groove or the wave
- The groove and the wave *evoke imagery* of body movements
- The groove and the wave are *recognisable*, recognised as being different or similar in certain ways to something we have already experienced.
- We are highly sensitive and responsive to the way fine-grained transformations over time in the groove or the wave are *transformed*. The changes make a difference to our experience
- The groove and the wave can be responded to, or be a basis for, *shared experience*. It makes sense to 'discuss' these nuances with other people, i.e. to exchange responses, to *share the experience*.

5.3.10. Non-symbolic aspects of meaning emerging from audio-visual interaction

With reference to the ballet *Rooster* I discussed how meaning emerges as a result of interactions between connotative meanings in music and movement, based on the understanding that music and movement constrain and contextualise each other reciprocally. This is what Cook refers to as *emergent meaning*, a feature that characterises musical multimedia. A similar understanding is implied in the expression *added value*, as proposed by Chion (Chion, 1994). In Cook's analysis of the Citroën commercial, mentioned in chapter 2, the emergent meaning is approached on the basis of how connotations of music and images interact, i.e. from a symbolic perspective.

The question is how meaning emerges as a result of the interaction of non-symbolic aspects. In this regards, the audio-visual analysis proposed by Chion, with his emphasis on "the wave as a dynamic model" rather than as a symbol (Chion, 1994; p 198), is an interesting account. As noted previously, Chion assigns audio-visual synch points a prominent role in the emergence of correspondences and in establishing a relation between sounds and images. In this regard, he claims that the temporal distribution of stronger and weaker synch points constitutes what he denotes to be the audio-visual rhythm of the film scene. This kind of rhythm may be seen as an underlying non-symbolic aspect of the narrative, a dramaturgical component. The audio-visual flow, or phrasing, the way the scene pushes forward towards a climax, pauses, the way tension is released, the way the scene is shaped by timing and dynamics, relates to ‘how’ the story is told, as opposed to the symbolic, psychological or semiotic aspect of the scene.

This non-symbolic level of meaning is of primary interest in Chion’s perspective of analysis, which means also that one of the main themes of analysis is the distribution and nature of synch points and the way these play a role in the narrative. For example, in the model analysis of the prologue scenes of Ingemar Bergman’s *Persona*[39], Chion points out that the overall “loose” audio-visual synchronisation that characterises the whole sequence is contrasted with the incident when a nail is hammered into a hand. In this camera shot, each blow with the hammer is brutally synchronised with the sound of the hammer hitting the nail. Chion makes the point that the asynchronous audio-visual relation supplies the sequence with its characteristic atmosphere; the “sequence derives its strangeness from the scarcity of synch points” (Chion, 1994; p 207). For example in a scene from a mortuary, we hear the dripping of some kind of liquid and the sound of footsteps, but the source of the sounds is in both cases not clearly given, and it is also not apparent how the sounds relate to the images. The ambiguous audio-visual montage evokes the spectator’s imagination, thus promoting added value (or emergent meaning in Cook’s terminology) and Chion concludes that the “treatment of these sounds of footsteps and dripping suggests a place where death is a common everyday phenomenon” (Chion, 1994; p 210).

The contrast between sequences of loose synchronisation and the scenes in which sounds and images are synchronised much more explicitly become meaningful suggestively and affectively; i.e. we are affected by the changes in synchronisation in the way our attention is changed and imagery is evoked. Furthermore, we may sense the concrete perceptual features of the mortuary; the smell and temperature of the room suggested in the sounds “of footsteps and dripping”.

With this in mind, I shall further discuss the way meanings may emerge from interacting non-symbolic aspects. Considering the presupposed prominence of goal-points discussed earlier, it is reasonable to believe that points of synchronisation, understood as *emergent goal-points* of audio-visual interaction, affect chunking and shaping.

The perception of the beginning and end of a musical phrase might in some cases, due to multisensory integration, be slightly altered when combined with a gestural unit; and vice versa, the boundaries of the gestural phrase may be temporally displaced accordingly. The role of interaction with respect to chunking is illustrated in the drawing in figure 22. The boxes drawn with full lines indicate the timing of boundaries in music and movement respectively when they are perceived separately. The dashed box suggests that when the beginning and the end of the auditory and the visual event occur sufficiently close to each other in time (although not exactly synchronised), the two events will appear to the perceiver as one fused audio-visual chunk. Due to multisensory integration the perceiver will interpret the auditory and the visual events to begin and end at the same point in time:

[39] Bergman, I. (1966). Persona [DVD]: Tartan.

Figure 22. The drawing suggests that the perceiver's interpretation of timing of the beginning and end may be changed due to multisensory integration, so that it appears that the auditory and the visual chunk begins simultaneously.

Also the weight of a goal-point will, as a result of multisensory integration, be strengthened (and in some cases weakened), when the goal-point of a musical phrase and a gestural chunk occurs at the same moment. This will affect the shape of the dynamical contour of a chunk; first, the emphasis of the goal-point may be increased and the temporal location may be slightly displaced, and second, the dynamical profiles of the prefix and the suffix may be stretched or compressed in time. This may lead to considerable changes, i.e. that dynamics evolve more gradually or more abruptly. And these changes in shaping and patterning may afford slightly different non-symbolic aspects of meaning in terms of affective-response characteristics. In the drawing in figure 23 I have tentatively indicated how an audio-visual dynamical shape (dashed curve) may differ for example from the dynamical contour of the musical phrase when heard alone (full line curve):

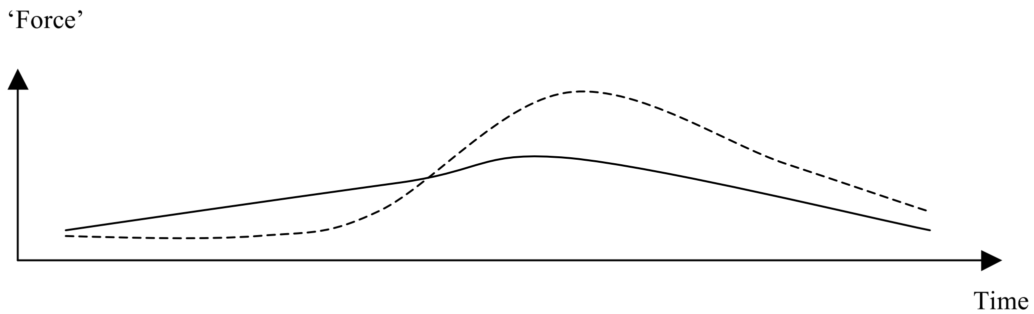


Figure 23. The illustration suggests that the dynamical shape of musical sound may be changed when combined with a visual element.

Pursuing this line of thinking, we may imagine a series of interacting musical and gestural events which involve sequences with a high degree of audio-visual convergence alternating with sequences with a lower degree of convergence, i.e. there is a fluctuation between 'loose' and 'tight' synchronisation. In other words we have a temporal unfolding in terms of convergence/divergence that is similar to the way sounds and images are combined in the opening sequence of *Persona*. This situation may be illustrated schematically as in figure 24:

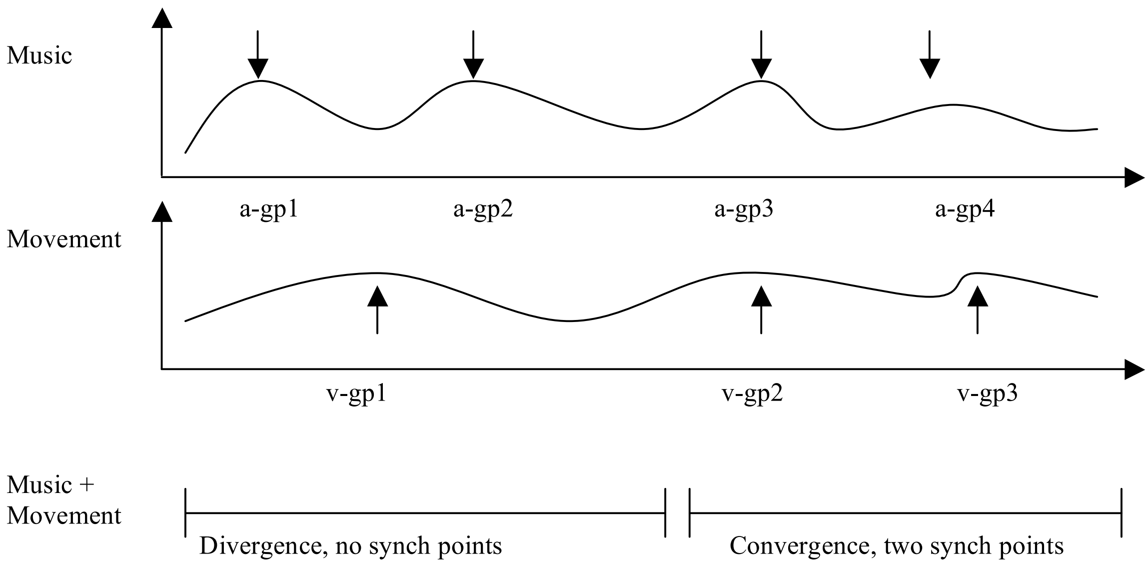


Figure 24. Two segments of an audio-visual combination in which the occurrence of synch points affects the sense of 'loose' and 'tight' synchronisation

In the first part the auditory goal-points (a-gp1 and a-gp2,) are not synchronised with the visual goal-point (v-gp1), so that no points of synchronisation emerge (a goal-point understood as a more weighted moment that the music/movement is 'heading towards', see discussion earlier in this chapter). In the second part, sonic and visual goal-points are sufficiently synchronised so that two synch points are perceived and the musical and gestural phrases are globally perceived as corresponding to each other. Thus from the interaction we have theoretically two sources that affect meaning in terms of non-symbolic aspects: firstly, we have a temporal ordering of contrasts in the fluctuation between divergence and convergence; secondly, since the weight of the synch points, due to multisensory integration, is changed compared to the emphasis of goal-points, for example in the musical phrases, we may assume that the contrasts between moments of articulation and moments of less tension (prefix and suffix) are changed accordingly. Consequently, and because of the change in contrasts, the interaction of the two media results in a rhythmical flow that is slightly different from the phrasing of the music and movement, respectively heard and seen in isolation. This is the phenomenon, or added value, that Chion addresses in the expression *audio-visual flow* (Chion, 1994).

The example also suggests that synch points affect the degree of perceived correspondence; i.e. that synch points serve the purpose of being some kind of glue that binds the two streams together. Again, I would propose that this may be understood in light of the perceived salience of goal-points in a peak structure; the synch point contributes to the perception of correspondence in the way the two streams are defined as one singular event.

This section has mainly focussed on the role of *goal-point synchronisation*. Additionally, a 'looser' kind of synchronisation may be observed when the auditory and visual events are synchronised in terms of chunks, i.e. a *window-synchronisation* that is based on the simultaneously perceived beginnings and endings of musical and gestural chunks.

5.4. Discussion: terminology and implications for analysis

Having kept the symbolic and the non-symbolic systematically apart throughout the chapter it seems timely to put them back together as two inter-related aspects of meaning. The theoretical considerations presented as well as the examples have in different ways suggested that the distinction is not clear-cut. I think that Laban's concept of effort, understood as a compound of physical features and an underlying "mental attitude" (see chapter 4), illustrates the point. Also McNeill's studies of co-verbal gestures point in this direction; on the one hand they are closely connected to the symbolic meaning of the speech they are accompanying, but at the same time they reflect non-symbolic aspects of thought (McNeill, 1992). For example, when telling the story of climbing up a steep hill, the slow, and somewhat strained, upward gesture of the hand depicts the dynamical shape of the event.

Furthermore, the blurred boundary has become apparent in the way I have approached the old woman's hand-gesture from many different perspectives of meaning. Similarly, with reference to the example from *Rooster*, it seems that the main gestural motif, the rooster strut, alludes to masculinity from two angles; first, by indirectly symbolising the walk of a cockerel, and second, more directly in the way the performance of the pattern, the dynamical-kinematical shaping, articulates masculine characteristics such as strength and the willingness to fight. The same two-sidedness is illustrated in a rhythmical groove: On the one hand the groove is responded to as a rhythmical, non-symbolic expression; on the other hand the specific groove signifies symbolic meanings within its socio-cultural context. In both cases, it seems that this blurred boundary is evoked in the way the form of the signs, i.e. the dynamical-kinematical shaping, is not arbitrary vis-à-vis their signification.

In my view, these observations and reflections converge in theories that understand meaning as being fundamentally embodied. As outlined at the end of chapter 3, Johnson has worked out a theory of meaning that views bodily experiences as a primary source of abstract thinking. For example, the metaphor 'balance' is embodied in the numerous and diverse experiences of balancing one's body (Johnson, 1987), thus suggesting a direct link between non-symbolic and symbolic aspects of meaning.

5.4.1. Terminology: *levels of meaning* vs *aspects of meaning*

Throughout the chapter I have consistently used the expressions *aspects of meaning* and avoided the quite commonly used *levels of meaning*. I have also preferred the term *non-symbolic* to *pre-symbolic*. These choices are based on theoretical considerations.

First, the term *level* may suggest that different levels of meanings are layered so that chunking precedes rhythmical patterning, which again precedes the interpretation of emotion and intentions, and that these levels, or categories of meaning, relate to each other in a linear (a 'bottom-up', or alternatively, 'top-down') manner. Likewise, the use of pre-symbolic would assert that something is perceived prior to the symbolic, so that the symbolic is based on the pre-symbolic. Thus the terms level and pre-symbolic suggest that perception is characterised as processing information through stages, as illustrated in figure 25:

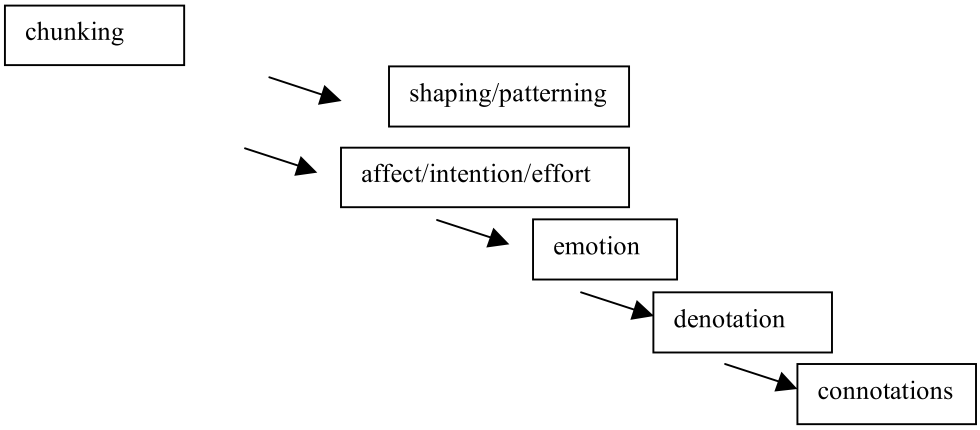


Figure 25. Emergence of meaning understood as a linear processing through stages.

This model in figure 25 is in line with the *information-processing paradigm*, i.e. that cognition is a matter of processing information that is already there in the environment. In this view the processing proceeds from the rudimentary levels of chunking to higher levels of cognition, and information is understood as something that is contained in the event independent of the perceiver. The perceptual levels of chunking/shaping/patterning are understood as steps on the way to cognising, which also suggests that perception/recognition are separated from cognition.

The understanding offered by ecological theory is quite different from this view. Perception and cognition are seen as interrelated processes, and characterised as the picking up of useful information. This means that perception/cognition are not about processing information through stages; the perceiver picks up aspects of the event which appear to be meaningful in that specific moment of awareness, and which are useful in terms of establishing a relation to the event. Understood in this manner, information is not present in the event, but is relative to the relation between the perceiver and the event, both contextualised by previous experience as well as by conditions provided by the present situation. To visualise this understanding, the components in figure 25 may be rearranged so that they are 'democratically' scattered in an experiential space, see figure 26:

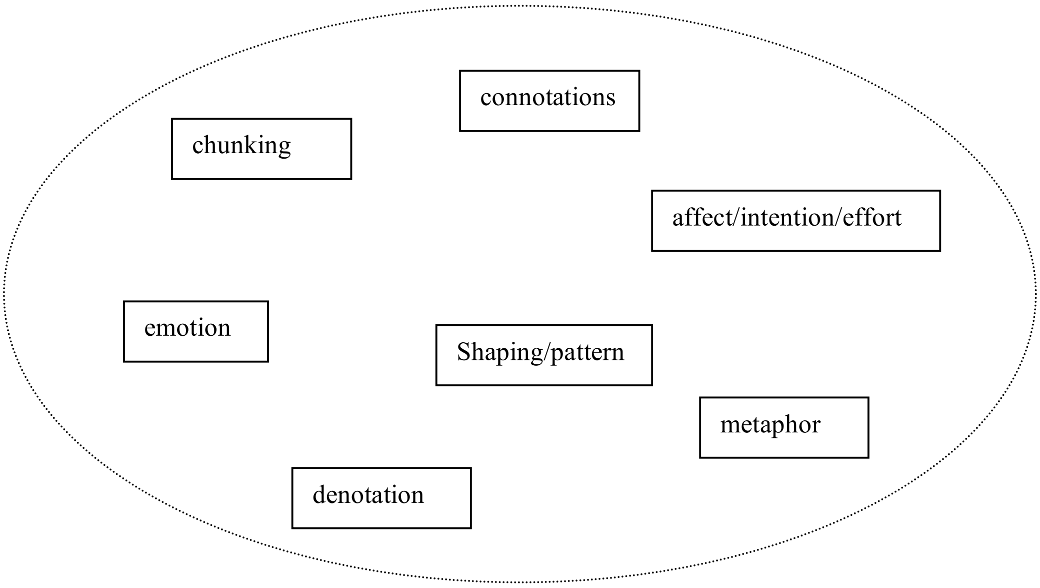


Figure 26. Emergence of meaning understood as an immediate picking up of any aspect of an event that attracts the perceiver.

In attending to events and trying to make sense of them meaning emerges on the basis of the way some of these aspects of meanings attract our attention and how they interact with each other. In the case of the old lady's hand-gesture many of us might be attracted and affected by the lightness, the effort quality of the gesture rather than by the semi-linguistic meaning of "waving goodbye"; which is not to say that the emblematic meaning does not play a part. Similarly, in *Rooster*, the rooster strut may immediately draw our attention to symbolic meanings related to conventional socio-culturally determined ideas of masculinity, but this connotation is most likely also intermingled with features of the particular performance, such as effort and shaping/patterning qualities. What this kind of understanding suggests is that different aspects of meaning do not relate to each other in a fixed manner, i.e. each aspect does not take a specific 'location' in the experiential space in relation to other aspects.

The experiencing of the particular event is characterised by the way we pick up what is useful or attractive to us at a particular moment; the aspect(s) that optimise understanding. This is where *affordance* comes in as an explanatory concept, i.e. what we pick up or utilise as useful information is constrained by what the event affords, and the affordances of an event are relative to the perceiver, the features of the event, and the relation between the perceiver and the event. In accordance with this line of thought, Godøy has proposed the expressions *multiple affordances* and the *multi-functionality of gesture* to refer to the understanding that the same gesture may carry different aspects of meaning or serve different functions at the same time (Godøy, 2008a).

5.4.2. Terminology: issues concerning the 'non'-structure

Based on these reflections the expression *pre-symbolic* seems to bear unfortunate connotations in the way it suggests a stage-like understanding of perception and cognition. For this reason I have chosen to replace pre-symbolic with *non-symbolic*. Accordingly, a variety of terms may be used: *non-conceptual*, *non-symbolic* or *non-verbal*. These terms may be appropriate for distinguishing the mode of experience, or aspect of meaning, from a conceptualised or symbolic perspective of meaning. The problem with this conceptualisation is that the 'non'-structure tells us what it is not, but does not directly convey the nature of this mode of experience; i.e. in what ways the wave as a dynamical form is meaningful to us. To navigate through the diversity of terms the following list may be helpful:

symbolic conceptual what verbal	⟷	pre-symbolic/non-symbolic pre-conceptual/non-conceptual how pre-verbal/non-verbal
conventional culture		individual nature
cognition indirect mediated distance	⟷	recognition direct affective/responsive/body-based presence

Table 4. The table gives an overview of terms related to the *symbolic* (to the left) and the *non-symbolic* (to the right).

These pairings all correspond to the symbolic/non-symbolic distinction in different ways. The first pairs (symbolic - pre-symbolic/non-symbolic; conceptual - pre-conceptual/non-conceptual; verbal – pre-verbal/non-verbal) address the form of articulation, whereas the conventional/individual/culture/nature divisions encompass the aspects of meaning in terms of constraints, i.e. that symbolic meanings are determined by social-cultural conventions. The lower part of the table suggests differences in the mode of awareness, or presence:

- It appears that non-symbolic meanings are characterised by the way the perceiver is present 'in' the event, as opposed to taking a more distant position.
- The perceiver may for example recognise a pattern or a shape, and responds to it directly without 'thinking' about how this pattern or shape may be labelled, for example in terms of a specific feeling such as sadness or joy (hence, the opposition between cognition/recognition).
- The terms *direct* and *affective*[40] allude to the immediacy that characterises non-symbolic meanings. The symbolic aspects seem to involve 'more thinking', whereas non-symbolic aspects are meaningful through their direct and immediate body responses, i.e. overt actions or covert motor imagery.

[40] Here *affective* refers to the way it is used by some researchers of emotion as denoting the immediate reaction to stimulation, see http://en.wikipedia.org/wiki/Affective

There are difficulties associated with all of these pairings. The 'pre'- and 'non'-structures in the upper part of the table have already been discussed. Furthermore, the middle part of the table implies that symbols rely on conventions or contexts, whereas pre-symbolic meanings are independent of contexts and culture, thus asserting a contextual/non-contextual division. This is also problematic because both symbols on the connotative level, or as non-linguistic signs, as well as non-symbolic meanings, are constrained by and sensitive to the contexts they appear in, only that the contextual source is different. In the case of symbols the context is constituted by historical and socio-cultural practices, whereas for non-symbolic meanings the context is provided by the history of the body/nature relation. Finally, it may be argued in light of ecological theory that the picking up of useful information is not more or less direct in the cases when symbolic meanings appear to be the most salient features of an event, as discussed in connection with *Rooster*. Or in other words, picking up useful information is always direct by nature, whether this information is attached to the symbolic or the non-symbolic.

This leaves us with the terms *affective*, *responsive* and *body-based*. In my view, these terms seen together have a high descriptive value in terms of capturing the nature of non-symbolic meanings. I would propose that non-symbolic meaning emerges on the basis of a closed affect/response cycle, which means that something in the event affects the perceiver so that he/she responds. The response, or better, the perceiver's awareness of the response supplies the event with meaning. Put differently, the event becomes meaningful through the way I respond to it and through the way I am aware of this response. I have attached the term *body-based* to this affect/response cycle to denote that the response may be grounded in body actions or covert motor imagery.

Another expression that has been introduced in music research is *sub-symbolic*. In a discussion of how to represent music, Leman distinguishes between three modes of representation (Leman, 1993):

- *acoustical representations*, characterised as iconic representations, such as waveforms, sonograms and spectrograms
- *symbolic representations*, characterised by a separation of form and content, as for example in musical notation
- *sub-symbolic representations*, characterised by our responses to musical sound.

Understood thus, the expression *sub-symbolic* seems to be analogous to the way *non-symbolic* is used in this thesis, i.e. in the way *sub-symbolic* is connected to the responsiveness of the perceiver:

> The property called responsive could ultimately lead to a criterion for distinguishing subsymbolic representations from other forms: In order to have a subsymbolic representation I suggest that there must be a causal way by which the particular description can be justified to come into existence. The information is not stored and then interpreted by the system (as in the symbolic approach). Rather, the memory *is* the interpreter: A stimulus gets its meaning in virtue of the response with respect to the environment. The memory is a resonator (Leman, 1993; p 134).

To round off this discussion one final comment should be added. The *conceptual* vs *non-conceptual* distinction could assert that 'more understanding' is involved in symbolic aspects of meaning. The ecological approach would claim that this is an unfortunate distinction because any aspect of perception and cognition is dependent on understanding. For this reason Alva Noë proposes that the conceptual vs non-conceptual distinction should be totally avoided (Noë, 2004), i.e. catching a ball in flight is as conceptual as knowing the semantic meaning of the word 'ball', only that the understanding in the first case is sensorimotor.

5.4.3. Implications for audio-visual analysis

A recurrent theme in this discussion has been that the boundary between the symbolic and the non-symbolic is not at all clear-cut, and that the different aspects of meaning influence each other reciprocally. In other words, the distinction between the 'what' and 'how' aspects of a movement or a musical phrase which was suggested at the beginning of this chapter seems difficult to uphold.

I have for example previously suggested that movements we do when listening to music, such as swaying or nodding 'in time' with the music, may be understood as reflections of the covert imagery of sound-producing movements, so that these movements as direct responses to the music should be understood as *non-symbolic*. The alternative interpretation would be that such movement-patterns articulate socio-cultural expectations, depending on the musical style and movement 'codes' agreed upon by a community of listeners (and dancers). Understood thus, the movement response and its relation to the music becomes symbolic. Clearly, these aspects interact so that the movement is grounded both in the symbolic and the non-symbolic.

Despite these considerations, I still think it useful, at least as far as theoretical and analytical reasons are concerned, to make a distinction between the symbolic and the non-symbolic. The analyses of audio-visual variants that are conducted later in the thesis are grounded in the idea that it is highly relevant to investigate music-movement correspondences in terms of non-symbolic aspects of meaning. This idea is also directly related to the empirical material of the thesis. The musical material used to collect the free dance-movements material was selected to direct the focus of attention and analysis towards non-symbolic aspects of meaning. Or put differently, the conditions provided by the music-movement relation afford correspondences on the basis of dynamics and kinematics in music and movement, rather than symbol-based correspondences.

This means that the approach to analysis is relative to the empirical material, which again means that the same approach would be less relevant to the analysis of music-movement relations in other performances of dance. The ballet *Rooster* has been discussed and it seems to me that this performance calls for an analysis of socio-culturally determined connotations, although it might also be interesting to examine further the interplay between aspects of meaning, e.g. the musical-gestural groove on a connotative level interacting with the groove as affective-responsive meaning.

In another example, the ballet performance *Brothers* [41], I would again suggest that one's attention alternates between different aspects of meaning. The ballet was choreographed as collaboration between two young male dancers and is performed by

[41] Parsons, D. (1992). The Parsons Dance Company [DVD]: Danmarks Radio. RM Arts.

them. The narrative context is the relation of two young boys; they play and fight, and there are moments of uncertainty and intimacy. Also apparent are the correspondences between the dance-movements and the music composed by Stravinsky[42]; there are periods and moments of close convergence between the music and movements, and there are sequences in which the audio-visual relation is less tight. With reference to Chion's analysis of the opening sequence of *Persona*, I agree that the correspondences in terms of non-symbolic aspects, the wave as dynamic form to borrow Chion's term, play an important role. However, in my view an analysis of this performance should address connotations that emanate from the narrative and the relation of two young boys, as well as the audio-visual flow that emerges from the alternation between loose and tight synchronicity.

In the next chapters, the collected empirical material, the sound-tracings and the free dance-movements will be discussed. The analyses will be conducted on the basis of the ideas and concepts elaborated on so far:

- Chunks of music and movement will serve as the primary zoom level for analysis
- The analysis will focus on the way features in music and movement change over time using the distinction between dynamical and kinematical shapes as a starting point
- The analyses will be guided by the overall theoretical view that a set of co-evolving features (shapes), for example in the musical component, may *afford* a diversity of movement responses

[42] Igor Stravinsky. *Concertino for 12 instruments*. Recorded by The Danish Concert Orchestra.

Chapter 6. Analysis of sound-tracings

6.1. Introduction

The purpose of this chapter is to analyse and discuss the empirical material that was collected in an initial sound-tracing study (Godøy et al., 2006a)[43]. As explained in chapter 1, nine people listened to two series of short sound excerpts (2-6 seconds of duration). In each excerpt they were asked to draw on a digital tablet the movement they thought best matched the sonic excerpt (for a detailed explanation of the procedure, see below). In figure 27 the display to the left shows nine different drawn interpretations of a single trumpet sound, whereas the display to the right shows a set of drawings corresponding to the sound of a cymbal being struck:

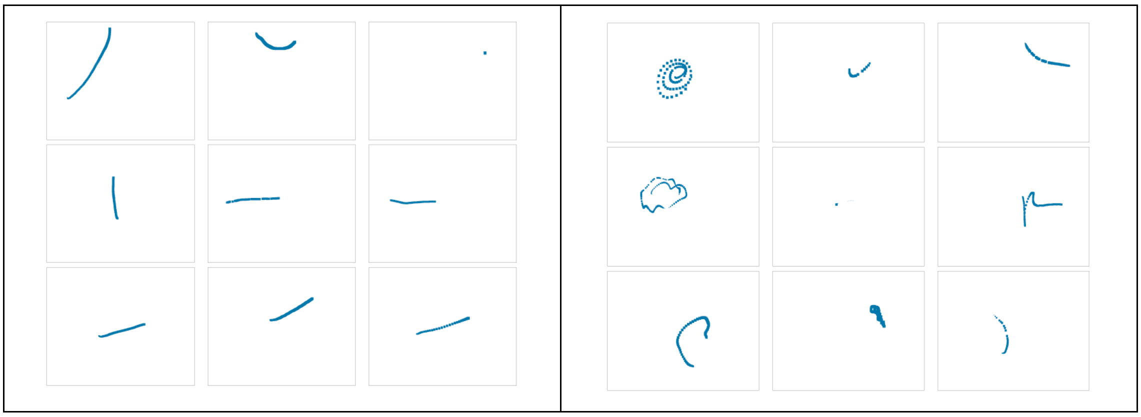


Figure 27. Two displays, each showing nine different drawn responses corresponding to two sound segments, respectively.

The reason for conducting the study was to examine motor imagery in sound perception. As explained in chapter 3, the core assumption in motor theory is that perception and action are intimately linked and that perception is constrained by the way we imagine, emulate or simulate the actions which are involved in a perceived event (Berthoz, 2000; Liberman & Mattingly, 1985; Wilson & Knöblich, 2005). The expression *sound-tracing* alludes to the understanding that an integral part of auditory perception is to trace motor processes in the sonic event, i.e. to explore changes in features in the sound by means of imagined (or actual) gestural responses. It is assumed that this imagery is based on salient features in the sound: the dynamic profile, pitch contour, or timbral/textural changes. Additionally, the imagery may be more directly related to the imagery of sound-production features. Thus we may distinguish between two complementing, competing or overlapping strategies

[43] The sound-tracing study was planned and carried out by Rolf Inge Godøy, Alexander Refsum Jensenius and the author, all participants in the Musical Gestures research group. I am indebted to this collaboration, both in terms of the work related to conducting the study itself and for the discussions we had during the work.

regarding motor imagery in sound perception, i.e. the contour strategy (shaping strategy) and the sound-production strategy. The example in figure 28 may be illustrative of the two general strategies:

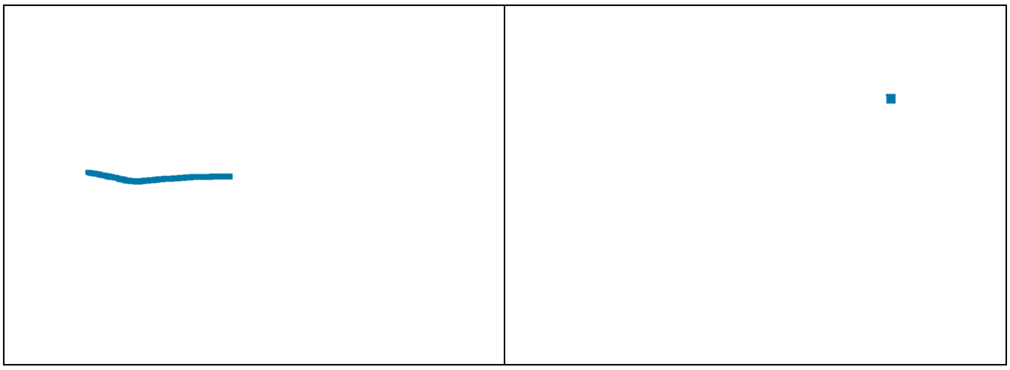


Figure 28. Two drawings interpreting a continuous trumpet tone. Respectively they employ a contour strategy- and a production strategy.

The corresponding sound in this case was a steady, non-changing (in terms of dynamics, pitch and timbre) trumpet tone. The drawing to the left appears to have been made on the basis of a contour strategy, i.e. the steady movement of the pen from left to right alludes to the non-changing feature of the sound. I would assume that the drawing to the right was made by pressing the pen on the tablet without changing position, and would suggest that the drawing alludes to the way the trumpet tone is produced by steady air pressure.

Provided we accept that the drawn response plausibly indicates the way the respondents imagine a movement of the body that matches a sonic event, we hypothesized that there would be similarities in the responses, i.e. we assumed that sound would afford movement in more or less the same manner across subjects. In other words, with reference to the main hypothesis of this thesis, the connection between sound and movement is not arbitrary; it is grounded in the cognitive and perceptual processes that were discussed in chapter 3. However, again referring to the main hypothesis, neither is this fixed, which means that we would expect differences between the drawings on one single display. With this in mind, the purpose of the study was to analyse the material with respect to the consistency in responses, i.e. *to what degree do people agree on correspondences between a drawing and a sound?* Additionally, the analysis aimed at describing the diversity in the material, for example with respect to the way different strategies may be employed in the *gestural exploration* of a sonic event.

The expression gestural exploration is central to this thesis. The drawn responses in the sound-tracing study (as well as the video-recorded dance movements to be discussed in chapters 7 and 8) are viewed as 'sketches' that reflect covert motor imagery. As such they are neither correct nor incorrect in terms of matching the sound, implying that they just reflect the way the individual sound-tracer was exploring a sound through a movement response. This kind of gestural exploration may be understood as a continuous (perceptual) process of proposing a hypothesis about the sound which is tested in the mode of performing a drawing[44].

In the first section below I shall explain in detail how the sound-tracing material was collected and comment briefly on consistency in the sound-tracing responses. Next, whether people agreed about the way the sound and drawing corresponded to each other will be further discussed on the basis of an interrater study. In this follow-up study, 20 people were asked to rate the match between the sound and the nine drawings. In the second part of the chapter, the results of the interrater study will be used as a starting point for a qualitative discussion of the way sound affords movement or vice versa, i.e. which features of movement are perceived as corresponding to features in a sonic event.

6.2. Collection of sound-tracings: design and procedure

Before proceeding to the interrater study and the qualitative analysis, I shall briefly describe the procedure used to collect the sound-tracings (see also Godøy et al., 2006a). We used two series of sounds. The first series consisted of 30 single sounds, i.e. sound segments that were not compounds of more than one tonal /timbral element (e.g. the sound of a trumpet or piano tone, the sound of a cymbal, or the sound of a voice uttering one single speech sound). Then the respondents (hereafter referred to as sound-tracers) were allowed a short break, before they were presented a second series of sounds. In this second series, 20 sound segments were chosen in order to introduce the sound-tracers to a higher degree of complexity, i.e. the segments consisted of more than one tonal/timbral element combined so that each segment was heard as one sonic gestalt.

A diversity of sounds was used ranging from electronic and traditional musical instruments to concrete, everyday sounds. Variation was obtained and systematised according to Pierre Schaeffer's typology of overall envelopes, i.e. impulsive sounds, continuous sounds or iterative sounds; and with respect to pitch and timbre features: stable, changing/unstable, or undefined (Godøy, 2006). This systematic selection of sound segments was carried out to facilitate/guide a later systematic analysis of sound/drawing correspondences., For example, a single trumpet tone would be characterised as 'continuous' with respect to overall envelope, and 'stable/unchanging' with respect to pitch and timbre/texture.

Instructions for the sound-tracers were carefully worked out. We wanted to instruct the participants so that they would associate the drawing with body movement, and not just to draw *per se*, as we had experienced in a pilot study that some participants in a few cases associated the sound with a drawing of the instrument itself. For this reason, we first explained that the purpose of the study was

[44] See chapter 3 where perception as hypothesis testing, a perspective proposed by Berthoz (Berthoz, 2000), is discussed. A similar view is found in Bregman's contribution on auditory perception (Bregman, 1990). He suggests the term *heuristic* to describe the process of making sense of a perceptual phenomenon i.e. as a functional response that may solve the "problem": "I use the word heuristic in its functional sense only, as a process that contributes to the solution of a problem" (Bregman, 1990; pp 32-33)

to analyse sound/movement relationships. Next, we alluded to movement in the following two phrases[45]:

- "In front of you have a tablet with a light grey surface. When you draw on the tablet with this pen the computer will record the movement you make on the tablet".
- "We would ask you […] to make the movement on the tablet that you think corresponds well to the sound. We would ask you to draw the sound the way you spontaneously perceive/experience it" (see appendix for full instructions in Norwegian and in English translation).

The tablet referred to above was a Wacom digital tablet[46]. The collected sensor data (obtained when the special pen touched the tablet) was stored digitally using a Max/MSP patch[47]. The patch was designed so that the participants got no visual feedback of their drawing so that they would concentrate on the movement rather than on the drawing as such.

Since the study aimed to examine reflections of motor imagery and the assumption that this motor imagery is immediately evoked in a non-symbolic manner when listening to music, we worked out the instructions to facilitate a response that was as immediate and spontaneous as possible. This was suggested in the last sentence referred to above.

For a similar reason, the Max/MSP-patch controlled the playback of the sound series so that a break of the same duration as the sound just played was inserted after the sound. The sound-tracers were asked to listen to the sound and immediately make the drawing in the break after the sound. In preparing the study, we considered letting the participants control the onset of each sound with a foot pedal or a computer key, but decided to proceed without interrupting the movements of the participants as we thought that other movements might disturb the spontaneity we sought.

The collected material consisting of data indicating time, pressure and location (X and Y coordinates on the tablet) was post-processed using Matlab-software and image-processing applications[48]. Each drawing was obtained by combining the location on the tablet (X,Y) with time (sample rate 100ms). The pressure data was included so that variations in pressure affected the thickness of the line, i.e. a movement performed with strong pressure would result in a slightly thicker line. The outcome was fifty displays each consisting of nine different drawings in a 3x3 matrix so that one display represented the drawings that corresponded to one single sound segment as drawn by the nine different sound-tracers (see example of displays in figure 29)[49].

---

[45] For the full text of the instructions in Norwegian and English translation please see Appendix.
[46] See: www.wacom.com
[47] The patch was created by Alexander Refsum Jensenius, see Jensenius, 2008, p 85 for details.
[48] A combination of *Max* and *GraphicConverter* was used for this purpose (Jensenius, 2008).
[49] The collected 3x3 matrixes of drawing responses have been combined with their corresponding sounds and are available as two 'audio-pictures'-files (created by Alexander Refsum Jensenius) on the CD, see the folder /sound-tracings chapter 6/.

The nine participants were selected so that people with different levels of expertise were represented: four of the sound-tracers were undergraduate or graduate music students, two were non-musicians but had had some musical training, and the remaining three had little or no musical training. For the study we included participants with different levels of expertise because we wanted to examine whether training made a difference on the sound-tracing responses. Considering the low number of participants, it is impossible to draw firm conclusions. Moreover, because most of the participants did have some musical training it turned out to be difficult to assign them to distinct categories of expertise. For these reasons, any conclusion regarding the role of training will have to be tentative.

6.2.1. Evaluation of consistency in responses

As described above, the raw material for further analysis was fifty displays with nine drawings on each display, each drawing representing the performance of one of the nine sound-tracers. A quick look through the material gives the immediate impression that there certainly are similarities between the participants in the way they responded to the same sound. Consider the two examples in figure 29:

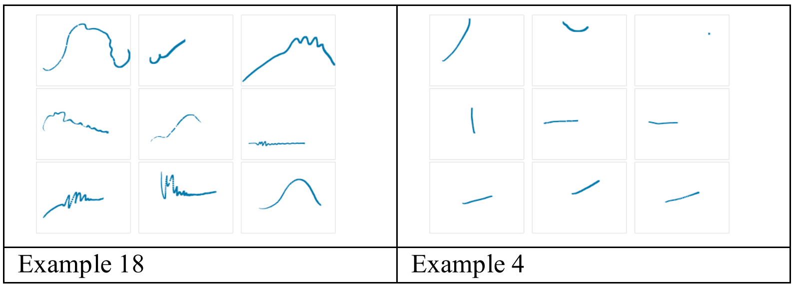


Figure 29. Two examples of displays, each representing nine different interpretations of two single sounds, respectively.

With a few exceptions the sound-tracers made their drawings with a movement from left to right. Accordingly, it seemed that the sound-tracers consistently imagined the timeline of the sound as moving from left to right. This means that similarities between the drawings should be considered with respect to the way the lines are drawn in this horizontal direction; and furthermore, correspondence between the way the sound and the drawing evolve over time should be viewed in this light.

In the example to the left in figure 29, four of the drawings have an arch-like contour, and two drawings have a contour that to some degree resembles this category. Three drawings do not have this predominant feature. Another shared feature is that in most of the drawings a part of the line is characterised by quick, scratchy curves. In the example to the right all the lines are evenly drawn and relatively straight. The exception is the drawing in the upper-left corner, which is one single dot. Five of the drawings have a horizontal trajectory and are about the same length. The trajectories of the other three straight lines differ from this. In conclusion, there are obvious similarities in the visual material presented here.

The preliminary discussion presented by Godøy and colleagues (Godøy et al., 2006a) concluded that the material demonstrated a fair amount of agreement in short segments with unambiguous envelopes and less agreement in fragments with more complex timbral/textural content. In light of the relative simplicity of the sound material, it was suggested that the consistency was not surprising. The participants in the study were all interviewed informally after the session, and many of them commented that the two-dimensional surface on which to make the response was quite limiting, so that they were forced to make choices. This suggests that their imagery of the sound was more vivid, e.g. more three-dimensional, than the actual response they were allowed to make. This problem increased as the complexity in the sound material was increased.

The preliminary suggestions about consistency were made on the basis of a run-through of the material intuitively and non-systematically. Therefore, to evaluate consistency with a more systematic approach, I carried out an interrater follow-up study. I asked a new group of respondents to judge which of the drawings on one single display best matched the corresponding sound example, and which drawings they thought least matched the sound. I assumed that some sound examples and the corresponding visual display would show a high degree of agreement between the raters, whereas in other cases the judgments would diverge more. It is reasonable to believe that the degree of interrater consistency may be understood in light of features in the sound and drawings so that thorough analyses of selected cases may reveal insights into how sound affords movement. Besides measuring interrater agreement thus, I also assumed that the results could be used for further qualitative analysis.

6.3. Interrater study

For the interrater experiment I selected 18 displays of the original 50. The selection was chosen on the basis of the following[50]:

- The material was ordered into two main groups, i.e. according to the sound characteristics of 'short sounds' and 'phrases'.
- These two groups were ordered into three subgroups according to my preliminary judgement of similarities/differences in the visual displays.
- A few examples had to be excluded due to technical problems in the drawings.
- Finally, nine examples from the two main groups were chosen and to obtain variation in the material, three examples were selected from each subgroup.

Having made this selection, the order of the drawings within the displays was randomly re-arranged so that in a series of nine displays each respondent would appear in all possible positions in the matrix, and also so that the respondents would not appear in a fixed order (e.g. N3 succeeding N2, succeeding N1 etc.). The re-ordering of the appearance of the displays was carried out to avoid the possibility that

[50] For the planning of this interrater study, I am indebted to Kjell-Ivar Øvergård and Cato Bjørkli at the Department of Psychology/University of Oslo for their advice. I would also like to thank the statistician Thorleif Lund for giving invaluable comments on the use of Kappa statistics and the rating procedure used in the interrater study.

the raters might 'learn' or get used to their judgements of match/mismatch appearing at specific positions on the displays. Last, the order of displays was re-arranged.

Twenty people were recruited to the study. Except for one above 40, the participants were between 20 – 35 years old. The majority were students of musicology at a bachelor/master's level, thus representing an intermediate level of musical training (e.g. 5-10 years of playing an instrument, but on a non-professional/semi-professional level). Two participants were trained dancers (both graduated from the National Academy of Ballet), and four had no formal musical education (but two of them reported a few years of playing/composing). Before the experiment they completed a form documenting their level of training.

The experiment was conducted as follows: one rater at a time, sitting at a table, was presented the series of displays one by one printed out on A4-paper arranged in order and stapled together into a booklet. The rater was informed about what the displays represented, i.e. that we had previously performed a study in which respondents had been asked to draw on a tablet the movement they intuitively felt matched a sonic event, and that the nine drawings on one display represented the response of nine different people to a single sound example. The sound was played back whenever the researcher double-clicked on the sound icon on a powerpoint slide, the sound emanating from two loudspeakers located on the same table. Then the rater was instructed to listen to the sound while looking at the corresponding display and asked to choose three drawings that he/she judged to match the sound best, and three drawings that he/she judged to match the sound least. The rater was instructed to label the best-match drawings with the number 1 (with a pencil directly on the paper) and the least-match drawings with the number 3. The three remaining drawings were not labelled but coded with the number 2 by the researcher after the session. The rater was allowed to listen to each sound example more than once while making his or her judgement before proceeding to the next visual display and sound example. Typically, judging one display took 30-60 seconds, and the entire session lasted between ten and fifteen minutes.[51]

Again, the wording of the instructions was intended to make the raters associate the drawings with body movement. Clearly, it is cannot be taken for granted that all raters did consistently view the drawings as representations of body movement. It should also be noted that no criterion for correspondence was specified. Such issues may have affected the proportion of agreement. This will be further discussed later in this section.

The responses obtained were entered into a data matrix in SPSS as exemplified in table 5:

[51] The sounds and drawing-displays used for the interrater study is available as a power-point file on the CD, see the folder /inter-rater examples chapter 6/.

Subject	Level	Drawing	Sound	R1	R2	etc…to R20
N3	E	1	1	2		
N1	M	2	1	3		
N2	M	3	1	2		
N5	E	4	1	1		
N4	N	5	1	2		
N6	M	6	1	1		
N9	M	7	1	1		
N8	N	8	1	3		
N7	E	9	1	3		

Table 5. The SPSS data matrix including the variables “subject” (the initial sound-tracers N1-N9), “level” (referring to level of musical training, i.e. E=expert, M=intermediate training, N=novice), “drawing” (numbered from left to right on each 3x3 display), “sound” (here referring to the first of 18 examples), and “R1” (referring to the first of 20 raters, each column containing the judgments of each drawing)

In the matrix the two first columns to the left specify the subjects N1-N9 who performed the drawings in the initial sound-tracing study and their level of expertise (E=expert, M=intermediate, N=novice). The next three columns show how raters (e.g. R1) judge drawing number 1 for sound 1, drawing number 2 for sound 1, etc. The judgements are coded from 1-3, according to the numbers used for best match (1) and least match (3), and the non-labelled drawings (2).

As a starting point for evaluating interrater agreement, the distribution of judgments has been shown using bar charts. Studying these charts reveals that for each example a few drawings stand out which are judged respectively as match and mismatch. For these drawings there is a high level of agreement among the raters. For the rest of the drawings there is a lower lever of agreement. In table 6 are two examples of bar charts showing the distribution of judgments:

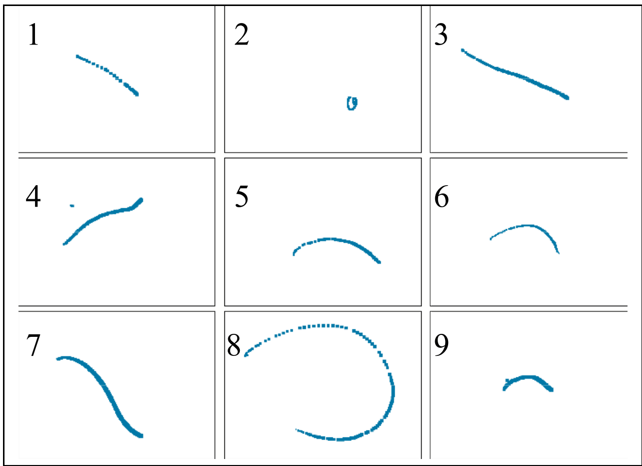
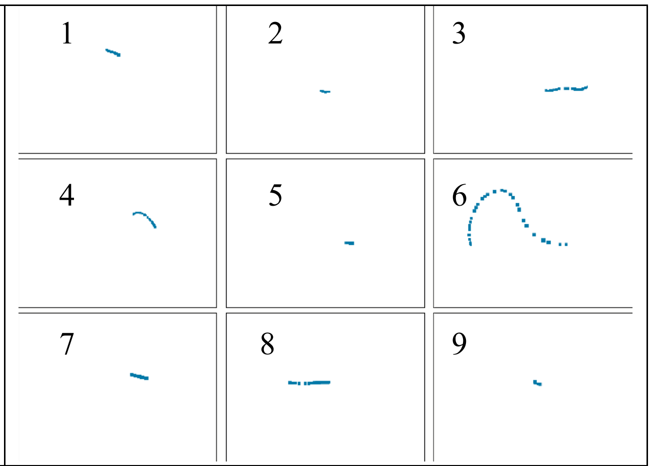
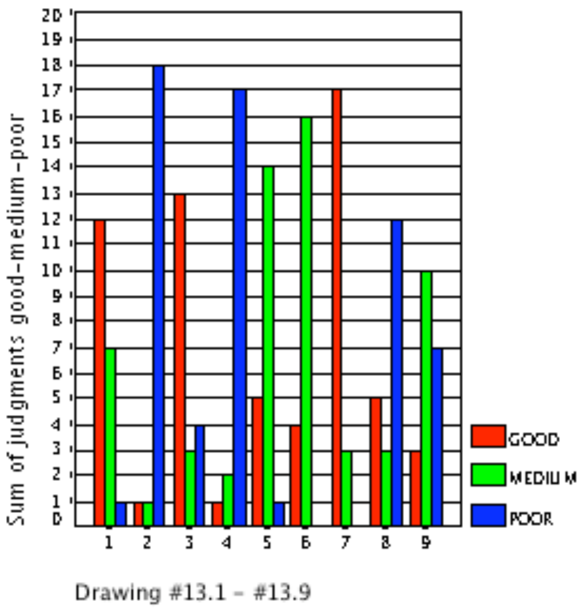
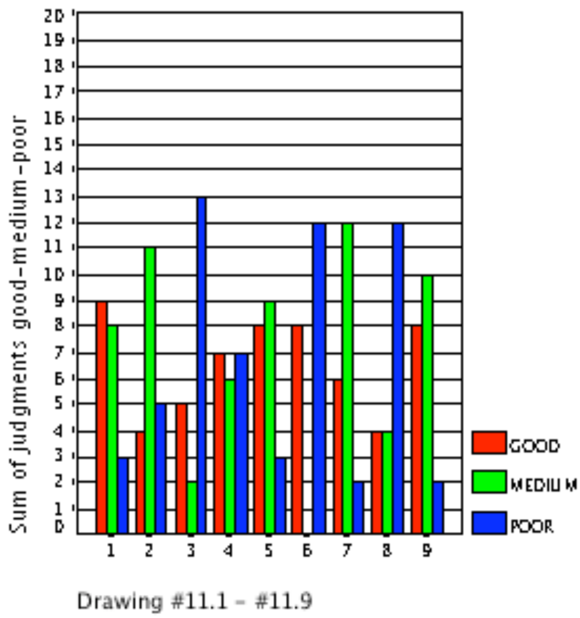
	
	
Example 13	Example 11

Table 6. The x-axis of the two bar charts represents drawing numbers 1 – 9 for the examples 13 and 11, i.e. drawing #13.1-#13.9 and #11.1-#11.9. The numbers of the drawings correspond to the order of the drawings in each display as shown (the drawing in the upper-left corner is labelled number 1 and the drawing in the lower-right corner is labelled number 9). The y-axis indicates the number of judgements for each category: The red bars indicate the sum of match ratings (good) for each drawing (1-9), and the blue bars indicate the sum of mismatch ratings (poor) for each drawing. The two charts show that examples differ with respect to concentration vs spread in the ratings.

From example 13 in the bar chart we can see that the drawings #13.7 and #13.2[52] respectively have been judged to match the sound well and poorly. However, in some examples the judgments are much more spread across all drawings, as illustrated in

[52] As explained in the legend for table, #13.2 refers to a drawing in example 13, i.e. number two from the left in the top row of the 3 x 3 matrix.

example 11. Thus we have some charts, as in example 13, in which single bars (either red or blue) stick out, indicating a relatively high degree of agreement on match/mismatch across the raters; and examples where each drawing has been judged more evenly. This means that the charts visualise to what extent the raters agree; i.e. the bar charts for examples 13 and 11 visualise differences in the level of agreement between the two examples. This difference between the examples will be further discussed later in this chapter, because the difference may reflect qualitative aspects of the sonic and visual material.

Agreement/disagreement between the raters may be further described by summarising responses in a confusion matrix for the pairs of raters. These show that the proportion of agreement differs from pair to pair. In table 7 I have illustrated this with a crosstabulation of the ratings of rater 15 (R15, music student) with three other raters: rater 12 (music student), rater 11 (music student), rater 3 (dancer), and rater 20 (novice, i.e. no formal music education, nor any training on an instrument). The diagonal with bold numbers shows the number of cases where the raters agreed on the categories 1 (best correspondence), 2 (medium correspondences) and 3 (least correspondence).

R15 vs R12.
Two music students, Kappa=0.392

Count

		R12			Total
		1	2	3	
R15	1	**34**	11	6	51
	2	13	**25**	13	51
	3	4	15	**32**	51
Total		51	51	51	153

R15 vs R11
Two music students, Kappa= 0.118

Count

		R11			Total
		1	2	3	
R15	1	**18**	19	14	51
	2	15	**22**	14	51
	3	18	10	**23**	51
Total		51	51	51	153

R15 vs R3
Music student vs dancer, Kappa = 0.363

Count

		R3			Total
		1	2	3	
R15	1	**28**	15	8	51
	2	18	**25**	8	51
	3	5	11	**35**	51
Total		51	51	51	153

R15 vs R20
Music student vs novice, Kappa = 0.324

Count

		R20			Total
		1	2	3	
R15	1	**30**	14	7	51
	2	15	**23**	13	51
	3	6	14	**31**	51
Total		51	51	51	153

Table 7. The table shows a comparison of Kappa values when one of the music students (R15) is paired with two other music students, one dancer and a novice.

I have chosen these pairs of raters to illustrate that music student R15 agrees at different levels with the two other music students R12 and R11. The matrices show that R15 agrees with R12 in 91 of 153 cases, and with R11 in 63 of 153 cases. For further comparison the table shows how the same rater, R15, agrees with a dancer and a novice.

Included in the table are also estimations of Cohen's Kappa, which is commonly used for the quantification of interrater agreement. Estimation of interrater agreement using Kappa statistics builds on the diagonal that describes the proportion of full agreement (bold numbers in the table above), and then corrects this for agreement that may have occurred by chance. In the R15 vs R12 case, the percentage of agreement is 59 %, and corrected for chance this produces a Kappa estimation of K = 0.392[53]. A value of 1 implies perfect agreement. Any value above 0 is considered to indicate agreement above the level of mere chance.[54]

There are diverging views about the interpretation of Kappa as a measurement of the level of agreement. Some would claim that any value below 0.70 indicates an unsatisfactory level of agreement, whereas others suggest an ordering of values, e.g. poor agreement = less than 0.20; fair agreement = 0.20 to 0.40; moderate agreement = 0.40 to 0.60 etc.[55]

Since Cohen's Kappa is only applicable for pairs of raters, I have measured Kappa for all possible rater pairs in the material. The results of the Kappa estimations are summarised in table 8 [56].

The Kappa estimations show that a majority of the values indicate some agreement above chance level, and that the values spread from very low (0.059, i.e. below chance level) to what is suggested above to be a 'fair agreement' level (around 0.40). In conclusion, the relatively low Kappa indexes indicate that there is a considerable proportion of disagreement between the raters.

The issue of diverging views about the way Cohen's Kappa may be interpreted in terms of the level of agreement will not be discussed any further. However, within the context of my material, the Kappa estimations quite clearly show that the observed level of agreement is generally low. In the following I shall use Kappa estimations with care as a tentative indication of the way agreement may differ because of different levels of expertise in the raters and because of differences in the sonic material (to be further discussed in the next section).

---

[53] See *Cohen's kappa. Index of inter-rater reliability.* http://www-class.unl.edu/psycrs/handcomp/hckappa.PDF for a description of the computational procedure used to estimate Cohen's Kappa. See Appendix for a numerical example

[54] http://en.wikipedia.org/wiki/Cohen%27s_kappa

[55] http://www.childrens-mercy.org/stats/definitions/kappa.htm. It should be noted that there is no general consensus about the appropriateness of Kappa statistics for interrater agreement. It is argued that the common statement that Kappa is a chance-corrected measure may be misleading, and that Kappa is not useful or reliable as a quantification of actual levels of agreement. Thus it is recommended to avoid ordering levels of agreement, e.g. poor vs fair agreement, on the basis of a Kappa measurement (for a discussion, see (Uebersax, 2006).

[56] Due to a technical problem concerning example 18, for raters R1 to R5 this example has not been included in the Kappa estimation, i.e. Kappa is estimated across 20 raters for 17 different examples, each example including 9 drawings.

	R1 (E)	R2 (D)	R3 (D)	R4 (E)	R5 (E)	R6 (E)	R7 (E)	R8 (N)	R9 (E)	R10 (E)	R11 (E)	R12 (E)	R13 (E)	R14 (E)	R15 (E)	R16 (N)	R17 (E)	R18 (N)	R19 (E)	R20 (N)
R1		.245	.265	.441	.284	.010	.314	.167	.108	.176	.010	.275	.343	.167	.373	.196	.314	-.010	.157	.167
R2	.245		.235	.275	.274	.108	.196	.225	.275	.108	.157	.255	.314	.147	.206	.167	.245	.078	.167	.265
R3	.265	.235		.392	.225	.000	.265	.176	.284	.137	.098	.216	.304	.196	.363	.137	.333	-.029	.265	.216
R4	.441	.275	.392		.382	.010	.225	.275	.245	.235	.118	.353	.333	.216	.265	.235	.422	-.039	.216	.235
R5	.284	.274	.225	.382		.147	.451	.255	.147	.147	.127	.324	.392	.206	.235	.265	.382	.000	.294	.255
R6	.010	.108	.000	.010	.147		.147	.137	.049	.118	.098	.108	.010	.128	.059	.020	.049	.039	.137	.127
R7	.314	.196	.265	.225	.451	.147		.216	.216	.206	.108	.196	.304	.255	.382	.255	.294	.157	.127	.216
R8	.167	.225	.176	.275	.255	.137	.216		.108	.157	.059	.108	.167	.127	.304	.216	.284	.118	.098	.245
R9	.108	.275	.284	.245	.147	.049	.216	.108		.147	.176	.147	.235	.127	.225	.186	.294	.049	.108	.118
R10	.176	.108	.137	.235	..147	.118	.206	.157	.147		.049	.324	.176	.147	.294	.216	.275	.088	.088	.167
R11	.010	.157	.098	.118	.127	.098	.108	.059	.176	.049		.029	.176	.196	.118	.010	.000	.069	.186	.029
R12	.275	.255	.216	.353	.324	.108	.196	.108	.147	.324	.029		.275	.265	.392	.284	.275	.049	.069	.284
R13	.343	.314	.304	.333	.392	.010	.304	.167	.235	.176	.176	.275		.196	.265	.216	.265	-.059	.225	.108
R14	.167	.147	.196	.216	.206	.128	.255	.127	.127	.147	.196	.265	.196		.225	.235	.137	.088	.157	.059
R15	.373	.206	.363	.265	.235	.059	.382	.304	.225	.294	.118	.392	.265	.225		.304	.275	.039	.147	.324
R16	.196	.167	.137	.235	.265	.020	.255	.216	.186	.216	.010	.284	.216	.235	.304		.225	-.010	.088	.216
R17	.314	.245	.333	.422	.382	.049	.294	.284	.294	.275	.000	.275	.265	.137	.275	.225		.049	.167	.275
R18	-.010	.078	-.029	-.039	.000	.039	.157	.118	.049	.088	.069	.049	-.059	.088	.039	-.010	.049		.000	.108
R19	.157	.167	.265	.216	.294	.137	.127	.098	.108	.088	.186	.069	.225	.157	.147	.088	.167	.000		.118
R20	.167	.265	.216	.235	.255	.127	.216	.245	.118	.167	.029	.284	.108	.059	.324	.216	.275	.108	.118	

Table 8. The table shows Kappa values for all pairings of raters. The raters are characterised as 'experts' (i.e. raters with an intermediate level of musical training, indicated by an E in the table), 'dancers' (i.e. raters trained as dancers, indicated by a D), and 'novices' (i.e. raters with no formal musical training, indicated by an N). It should be noted that two of the novices, R8 and R16 did have some musical background.

6.3.1. Discussion: Sources for agreement/disagreement

In this interrater study the evaluation of sound/drawing correspondences were measured on an *ordinal categorical* scale. The data are categorical since they have been sorted into the categories best, medium and least. The scale is ordinal because the judgments have been sorted in terms of the degree of correspondence using a scale that indicates only the ranking order, and not the 'distance' between each category (as opposed to an *interval* scale) (Pedhazur & Schmelkin, 1991). This means that the distance between best and medium may be different from the interval between medium and least, as illustrated in figure 30:

Figure 30. An *ordinal categorical* scale indicates the ranking order of different categories. The distance between the categories is not fixed as in an *interval scale*.

Two main sources are usually thought to contribute to agreement/disagreement in ratings on an *ordinal categorical* scale used here. The first source for disagreement may be found in the definition of the trait, i.e. whether the raters are employing the same definition of 'best correspondence'. The second source relates to the use of the scale, i.e. whether the raters differ in the way they draw boundaries between categories (Uebersax, 2006).

The instructions were given open-endedly; the raters were asked merely for their own judgment of best and least, and no criteria for correspondence were specified. Furthermore, the measurement is affected by the way the instructions forced the raters to choose three drawings as best and three drawings as least. This was because we wanted an spontaneous judgment of each display with nine drawings, and assumed that a more constrained 3+3 judgment with a focus on best and least would be quicker to perform than if the raters had to look carefully at each drawing and were free to decide how many drawings to assign to each category. However, this format has obvious disadvantages in terms of the reliability of measurements since it may force the respondent to use slightly different scales/different definitions of the trait for judging one set of drawings compared to another. In extreme cases the situation may appear thus:

- In one example, some of the raters may have found that most of the drawings corresponded well to the sound so that it was difficult to choose three drawings for best match. This means that some of the drawings that otherwise might have been rated as best had to be rated as medium (or even least).
- In another example, some of the raters may have found that most of the ratings corresponded poorly to the sound so that it was difficult to pick out any of the drawings for best match. This means that drawings that, according to the rater's

definition of correspondence, otherwise might have been rated as medium match are forced into the best category.

In fact, many of the raters commented that in some of the examples the drawings were either too similar, or they could not find any good match between the sound and the drawings so that they were forced to make their choice more or less randomly rather than using consistent criteria for correspondence. The 3+3 rating format implies that the criteria/scale may vary slightly, and in some cases considerably, between the examples.

In light of the measurement problems inherent to the chosen rating procedure, an alternative procedure might have been considered. For example, one option could have been to instruct the raters to make judgements on the basis of a 5-point scale with each step 1-2-3-4-5 indicating an increasing level of correspondence. This approach would imply each individual drawing being rated with respect to the level of correspondence with the sound, and according to an *interval scale*. The procedure has advantages compared to the 3+3 rating format since the rater is not forced to choose a fixed number of drawings as best match, but is in principle allowed to judge each drawing by the same criteria and the same scale for all sound/drawing combinations. Thus, we might expect a more consistent and 'truer' rating.

However, I believe that using this kind of rating procedure for judging sound/drawing correspondences would also have certain limitations:

- First, it is questionable whether it is in fact possible to clearly specify different levels for the sound/drawings to enable judgements according to an interval scale, i.e. where each level is clearly defined and the 'interval' between each level is the same. I suspect that this would lead the raters to make the judgements according to ranking order (e.g. "drawing A has a slightly better match than drawing B, so A is level 4 and B is level 3"), so that we in fact end up with an *ordinal scale*.
- Second, the raters would most likely have spent much more time judging one example of nine drawings vs one sound since they would have to attend more carefully to each drawing one-by-one. It would have been difficult to rate the same high number of drawings (a total of 162 drawings in the 3+3 rating procedure) and I assume that more time spent might have affected consistency in the use of scale and criteria from the beginning to the end of the rating session.

This suggests that a rating scale procedure would require more detailed instructions in terms of specifying criteria. The judgments would be more directed towards specific features so that the responses would be less immediate and less individual. Despite the limitations of the 3+3 format, the procedure was chosen since we wanted simple and straightforward instructions, an immediate response (not too much thinking), a relatively high number of drawings to be judged within a short time, and finally, because we did not want to specify criteria but to leave it up to the raters in each example.

Having commented on the choice of rating procedure, I shall now return to the discussion of sources of agreement/disagreement. Apart from the way these design-related issues may affect the observation of agreement/disagreement, it was not unexpected that the raters would make diverging judgments about sound-movement relations. On the one hand, I would suggest that the raters may have employed their shared ecological knowledge of sound-movement relations so that they used shared criteria to judge correspondence. Understood thus, the observed agreement reflects a certain robustness, or a *non-arbitrariness*, to sound-movement relations. Conversely, applying the hypothesis that sound-movement correspondences are perceived *flexibly*, we would in some examples expect the raters to use different criteria for the judgment of the same sound/drawing relationship.

A further source of agreement/disagreement may be found in the sonic and visual material, i.e. that some sounds afford a more precise and non-ambiguous movement response so that the sound/drawing correspondence is easier to agree on; or alternatively, that other sounds are less readily visualised in the restricted format which involves drawing on a tablet, or that the sound affords multiple options for a drawing response so that it is more difficult to find clear and unambiguous criteria for correspondence.

The qualitative analysis in the second part of this chapter has been organised so that examples of sounds/drawings are arranged in five groups with respect to the characteristics of the sounds. For example, in one group of the examples the sounds are characterised by a clear pitch contour, whereas in another group of examples the sounds have all a non-changing pitch feature. To illustrate this point, table 9 compares the Kappa measurements of two raters, R15 and R12, for two groups of examples. We can see that these two raters demonstrate a slightly higher level of agreement in examples with a clear pitch contour than in examples with a non-changing pitch. If we make the same comparison for raters R15 and R20, we can see that the level of agreement is considerably higher for ‘clear pitch contour’ than for the ‘non-changing pitch’ examples.

R15 vs R12
Agreement on ‘non-changing pitch examples’
Kappa = 0.250

Count

		R12			Total
		1	2	3	
R15	1	**5**	3	4	12
	2	5	**6**	1	12
	3	2	3	**7**	12
Total		12	12	12	36

R15 vs R20
Agreement on ‘non-changing pitch examples’
Kappa = .042

Count

		R20			Total
		1	2	3	
R15	1	**5**	3	4	12
	2	4	**4**	4	12
	3	3	5	**4**	12
Total		12	12	12	36

R15 vs R12
Agreement on ‘clear pitch contour examples’
Kappa = 0.333

Count

		R12			Total
		1	2	3	
R15	1	**13**	4	1	18
	2	4	**7**	7	18
	3	1	7	**10**	18
Total		18	18	18	54

R15 vs R20
Agreement on ‘clear pitch contour examples’
Kappa = 0.528

Count

		R20			Total
		1	2	3	
R15	1	**12**	4	2	18
	2	5	**11**	2	18
	3	1	3	**14**	18
Total		18	18	18	54

Table 9. Estimations of Kappa suggest differences in agreement in examples with clear pitch contours compared to examples with non-changing (or undefined) pitch.

Clearly, the number of cases is relatively low for each group of examples, especially for the non-changing pitch group, so the comparisons are shown for illustrative purposes only. Any further examination of the material in this regard would also have to consider estimations of agreement for a higher number of rater pairs to evaluate whether a general tendency could be revealed. Such differences in agreement on the basis of features of the sounds and drawings will be further discussed on a qualitative basis in the second part of the chapter.

The participants’ level of training, both in the initial collection of sound-tracings and in the interrater study, may also have contributed to the agreement/disagreement observed. It seems reasonable to assume that trained participants have acquired a more precise idea of sound-movement relations, so that they more consistently employ criteria for correspondence. This means that as a participant in the sound-tracing study the expert would visualise the sound more clearly and less ambiguously compared to the non-expert, which again would make it easier for the raters to agree on match-mismatch. Likewise, we might expect that the expert as a rater would judge the drawings more consistently.

Additionally, there is the question of the way the participants experience the connection between sounds and the drawings; whether it is regarded as strange so that the task of rating sound/drawings may appear unusual and difficult. We might assume that trained musicians are more used to describing music by 'translating' musical sound into other modes of articulation, so that the tasks in the collection of sound-tracings as well as in the interrater study appear less unusual and strange. This would again lead to more consistent criteria for correspondence.

As previously explained, mainly music students at bachelor/master's level participated in the interrater study. Additionally, two trained dancers and four raters without any formal musical training were included in the study. If we compare the Kappa values of the novices and music students we find no systematic differences. Except for one of the raters who had no formal musical education or training, who had Kappa values around zero, the other three novices agree with the other raters as to the similar level of agreement. The same applies to the two dancers. In fact, two of the music students differ from the others with quite low Kappa values (R6 and R11), thus this specific material suggests that the level and nature of musical training/experience do not affect the judgment of sound/drawing correspondences systematically.

However, if we take a closer look at the expert and non-expert drawings and how they are judged it appears that training may have some influence on the ratings. By separating the drawings made by the trained and non-trained sound-tracers in the initial study, we may compare interrater agreement with respect to the drawings made by experts with agreement on the drawings made by novices. Provided that we assume that trained sound-tracers exhibit a greater level of precision and employ criteria for correspondence more consistently, we might expect the material to appear less ambiguous to the raters, so that the level of agreement increases.

Accordingly, I estimated the agreement of raters R15 vs R12 and R15 vs R20 respectively, and compared judgments of 'trained' vs 'non-trained' drawings[57].

[57] The reader may have observed that in these matrices the row and column totals diverge, whereas they are always the same in the previous examples. The reason for this is that when we compute Kappa for only the 'trained drawings' for example, there is no longer an even distribution between the rating categories.

R15 vs R12
Agreement on ‘trained drawings’
Kappa = 0.474

Count

		R12			Total
		1	2	3	
R15	1	**23**	6	2	31
	2	9	**16**	4	29
	3		3	**9**	12
Total		32	25	15	72

R15 vs R20
Agreement on ‘trained drawings’
Kappa = 0.298

Count

		R20			Total
		1	2	3	
R15	1	**19**	8	4	31
	2	7	**13**	9	29
	3		5	**7**	12
Total		26	26	20	72

R15 vs R12
Agreement on ‘non-trained drawings’
Kappa = 0.226

Count

		R12			Total
		1	2	3	
R15	1	**2**	1	2	5
	2	1	**5**	5	11
	3	1	6	**13**	20
Total		4	12	20	36

R15 vs R20
Agreement on ‘non-trained drawings’
Kappa = 0.289

Count

		R20			Total
		1	2	3	
R15	1	**2**	2	1	5
	2	4	**5**	2	11
	3	3	4	**13**	20
Total		9	11	16	36

Table 10. The Kappa value is higher when the two raters R15 and R12 judge drawings made by trained sound-tracers.

This revealed that the level of agreement is markedly higher when the rater pair R15 and R12 judge drawings made by trained sound-tracers, i.e. Kappa = 0.474 (trained drawings) vs Kappa = 0.260 (non-trained drawings). For the rater pair R15/R20 there is however no marked difference.

In sum, the material suggests that the raters’ level of training does not affect the proportion of agreement. In the case of the R15/R12 pair however, it seems that the level of training may have influenced the performance of the participants in the initial sound-tracing study, so that the two raters agree on a higher level for ‘expert’ than for ‘non-expert’ drawings.

Again, it should be noted that the difference in the Kappa values in table 10 above is shown for illustrative purposes only, tentatively pointing in the direction that drawings made by ‘experts’ may be more consistently rated. A further examination of this would have to consider the Kappa estimations of all possible pairings of the raters. However, this issue will not be followed up here, since I believe that the material has additional weaknesses in this regard. First, the number of raters is relatively low, i.e. we would need more participants in each group, especially in the non-expert group. Second, it seems difficult to distinguish clearly between levels of training. Where do we draw the line between different levels of expertise: between the professional vs the intermediate, and the intermediate vs the novice? And what kind of training is relevant; piano playing or dancing, violin playing or music listening?

On the basis of the presented material, it is difficult to evaluate decisively to what degree the level of training affects the proportion of agreement/disagreement observed in the interrater study.

To sum up, the following sources for agreement/disagreement have been discussed:

- Sources related to the measurement of the agreement (categorical ordinal scale and the 3+3 format)
- Sources related to the phenomenon as such, i.e. that we may expect a certain proportion of agreement due to shared ecological knowledge about sound-movement relations, and a certain proportion of disagreement in light of the flexibility of sound-movement relations
- Sources related to sound characteristics, i.e. that some sounds are more readily judged in terms of correspondence with a drawing
- Sources related to the level of expertise

6.4. Qualitative analysis of criteria for sound/drawing correspondences

From this discussion of interrater agreement and sources of agreement/disagreement I shall now proceed to focus on individual examples of sound/drawing relations. The main purpose of the qualitative analyses is to discuss the criteria that the raters may have used to judge correspondence between sounds and drawings.

The starting point will be the agreement/disagreement that was observed in the interrater study. This implies that for one set of drawings (i.e. a display of nine drawings that 'belongs' to one single sound) that demonstrates a relative high degree of agreement I will assume that the particular visual material and the corresponding sound offer the possibility of employing relatively non-ambiguous criteria. Likewise, for single drawings that most of the raters agree have a good correspondence with the sound, I will assume that these drawings fulfil the criteria in a relatively precise manner, and that other drawings on the same display were not chosen either because they are less precise, or because the rater is not able to find any suitable criteria for a sound/drawing connection. The underlying premise of the discussion is that observed interrater agreement is used to indicate that some criteria for correspondence are "stronger" or more readily picked up on than others.

As explained previously, bar charts are useful to illustrate graphically how the raters judged each drawing. The chart in figure 31 shows the distribution of ratings for a trumpet sound (example 4):

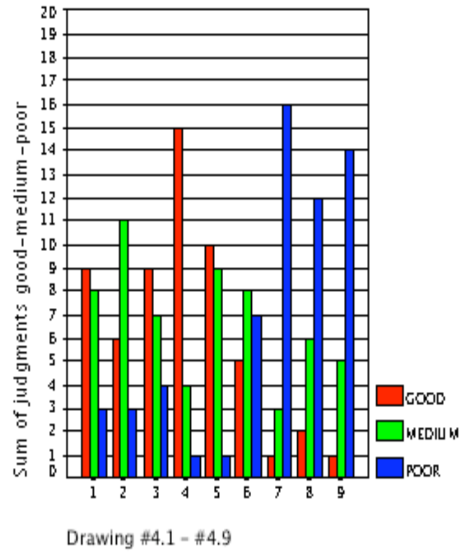
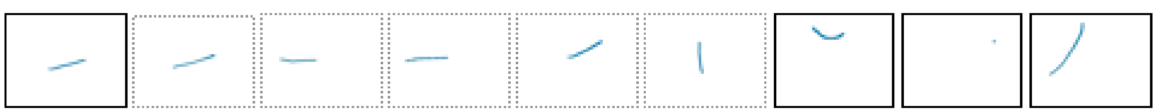


Figure 31. The ratings of drawings vs trumpet sound show that drawing #4.4 (x-axis) is rated the best match by 15 of the raters, whereas drawing #4.7 is rated the least match by 16 of the raters.

The chart shows that drawing #4.4 is rated best match by 15 of the 20 raters. I would consider this to be a relatively high level of agreement for sound/drawing match. For drawing #4.5 the agreement for match is lower. On the other end of the scale, drawing #4.7 is rated least match by 16 of 20 raters, which gives a relatively high level of agreement on mismatch. Based on the chart I have re-arranged the nine drawings for each example into a series of drawings, as shown in figure 32:



‘14’ ‘9’ ‘6’ ‘5’ ‘3’ ‘-2’ ‘-10’ ‘-13’ ‘-15’

Figure 32. The drawings are arranged in a series so that those rated best match are located to the left and the least match drawings to the right. The ordering is based on a score obtained after a re-coding. The drawings with a score above +10 and below -10 are marked with a full line to indicate that they are relatively consistently judged to be either best or least match.

The drawings to the left in the series are rated as match and the drawings to the right are rated as mismatch. The ordering of the drawings is based on the following:

- The rating categories are re-coded so that best= 1, medium= 0 and least= -1
- A score for each drawing is obtained by subtracting the number of least from the number of best drawings
- The drawings are ordered by rank with the highest scores to the left and the lowest scores to the right, as shown above

For drawings with a high level of agreement on match, the score will be close to +20, whereas the score will be close to –20 for a high level of agreement on mismatch. In some examples the judgment will spread so that most of the scores will cluster around zero. In such cases there is a relatively low level of agreement on all drawings.

For further illustration I have marked the drawings that show a relatively higher level of agreement on match/mismatch with a full line. I have chosen to mark the drawings with a score higher than and equal to +10 and those with a score lower than and equal to -10. The remaining drawings (with a score between -9 and + 9) are marked with a dotted line. For some of the examples, 3-4 of the drawings will be marked with a full line, such as in the example of the trumpet sound/drawing shown here; however, as will be shown later, in other examples none of the drawings is marked with a full line. Thus, this way of marking the drawings visualises differences in the level of agreement between examples. This means that the thresholds of +10 and -10 have been chosen to distinguish between examples with a generally higher level of agreement compared to examples with a generally lower level of agreement. This threshold has been chosen on the basis of my own qualitative judgment, as a reasonable level of agreement considering the flexibility of judgments that we may expect in rating sound-movement relations.

Ordering the drawings from match to mismatch is used to facilitate the qualitative analysis, which will focus on the criteria that the raters may have utilised in their ratings. On a theoretical basis, and with reference to the preceding chapters, I would suggest that the criteria listed below for correspondence were involved in the judgments. In the analyses I shall describe how these criteria are involved in sound/drawing correspondences and how they may overlap and interact:

- *Does the drawing correspond to contours of change in sonic features?*

That is, we may assume that in some cases the raters pick up on the contours of change in the sound, such as changes in pitch, timbre or dynamics, and that these are compared to shapes, contours or patterns of change in the drawings.

- *Does the drawing visualise imagery of the way the sound may have been produced?*

This criterion applies to imagery of the way energy/force is distributed in sound-production (dynamics of sound-production), i.e. a distinction will be made between impulsive, continuous and iterative sound-production. The question is whether the drawings somehow describe/sketch out the movements that may have been involved in sound-production.

- *Does the drawing correspond to amodal attributes that are apparent in the sound?*

This means that I will be aware of general amodal attributes (see chapter 3) such as number, duration, intensity and rhythm (i.e. understood as patterns).

With reference to the observed agreement/disagreement, I shall address two basic questions:

- In cases with a relatively high level of agreement, what do the raters agree on?
- In cases with a relatively low level of agreement, what do the raters disagree on?

I have organised the analytical procedure so that examples, which are similar in terms of sonic features, are discussed in relation to each other. Hence, five groups of examples are analysed: (1) one single tone that is characterised as non-changing, or undefined, in terms of pitch and timbral features; (2) one single tone with an apparent pitch contour; (3) successive sonic elements that appear as a pattern; (4) iterative sound-production; and (5) changes in dynamics and/or timbre. Initially, this way of organising the material demonstrates similarities in the drawn responses within each group of examples on an intuitive 'first –sight' basis. Moreover, the arrangement enables within-group, as well as between-group comparisons.

6.4.1. Group 1. Four examples with one single tone characterised by being non-changing (or undefined) in terms of pitch and timbral features

First, I shall discuss a group of examples in which the sound element is characterised by its non-changing features in terms of pitch and timbre, i.e. example 3 (one single piano tone)[58], example 4 (one single trumpet tone)[59] example 11 (speech sound, "meh")[60], and example 16 (crash cymbal) [61]. The visual responses to these sounds are shown in table 11. Included in the table are also bar charts summarising the ratings of each drawing.

In example 4 the sound segment is a sustained trumpet sound, i.e. there are no apparent changes in pitch, timbre or dynamics. Similarly, the piano sound in example 3 is characterised by a non-changing pitch and timbre except for the subtle timbral and dynamic changes that make it sound like a piano tone, i.e. a clear attack when the string is hit and a decay lasting for a total of 3.5 seconds. The segments differ with respect to sound-production because the trumpet tone is produced with a sustained force (air pressure), whereas the piano tone is produced with a more momentary force.

[58] Sample by Alexander Refsum Jensenius
[59] From general sample bank
[60] From Pierre Schaeffer's *Solfege de l'Objet de Sonore*
[61] From general sample bank

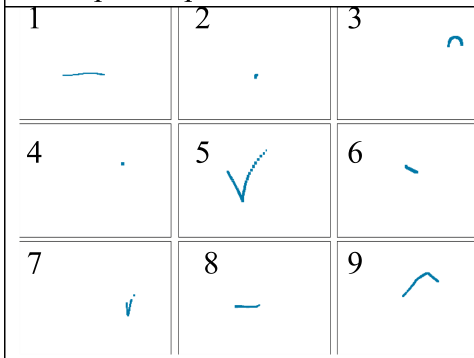
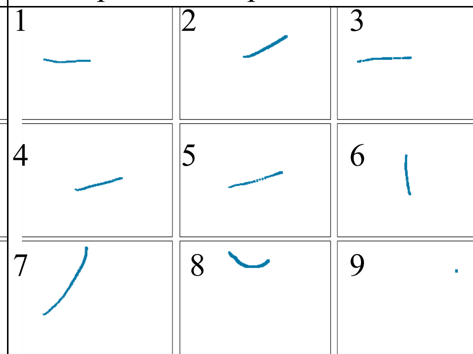
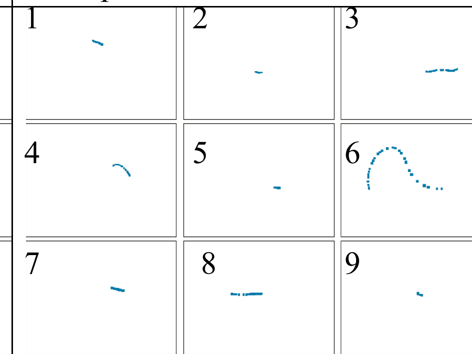
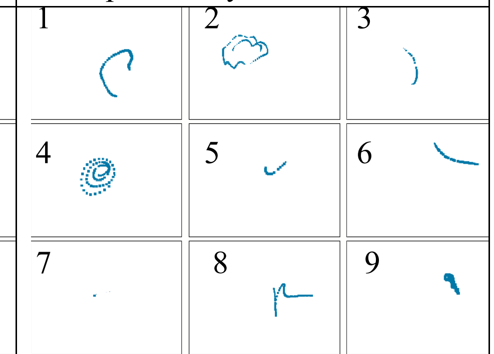
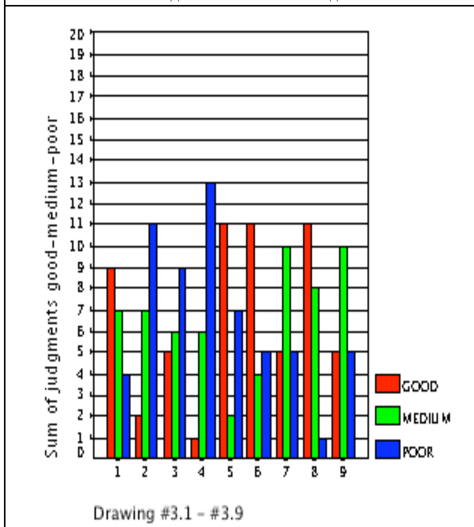
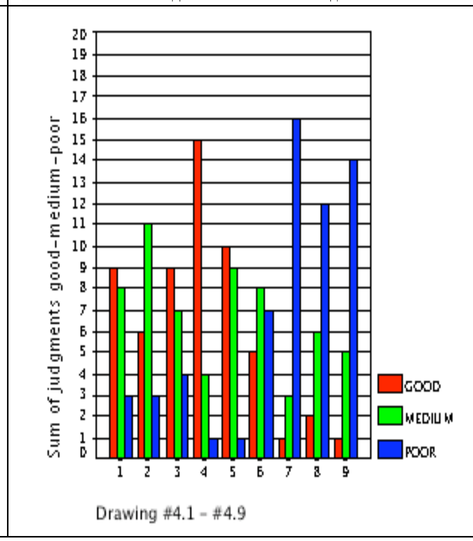
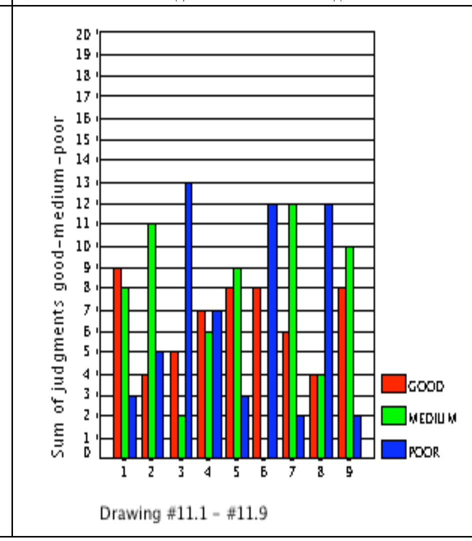
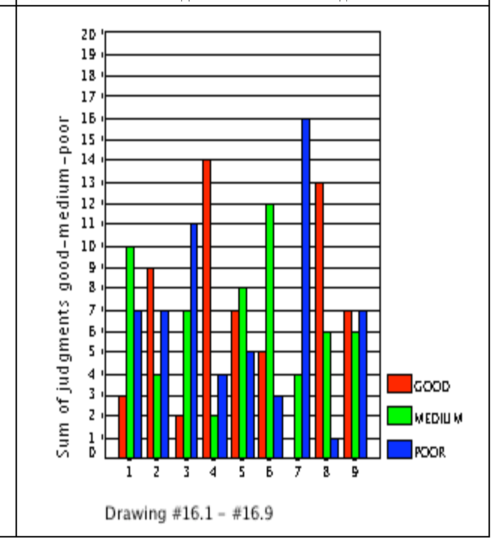
Example 3 - piano	Example 4 - trumpet	Example 11 – 'meh'	Example 16 - cymbal
			
			

Table 11. The table shows the drawn responses to the sounds in examples 3, 4, 11 and 16 together with bar charts summarising the ratings for each drawing.

In example 4 it appears that the raters favour a straight, unbroken, and relatively horizontal line. The curvy lines in drawings #4.7 and #4.8 as well as the dot in drawing #4.9 are generally agreed not to fit well with the trumpet sound. Of the remaining straight lines, drawings #4.4, #4.5 and #4.1 obtain the highest rating with respect to good correspondence.

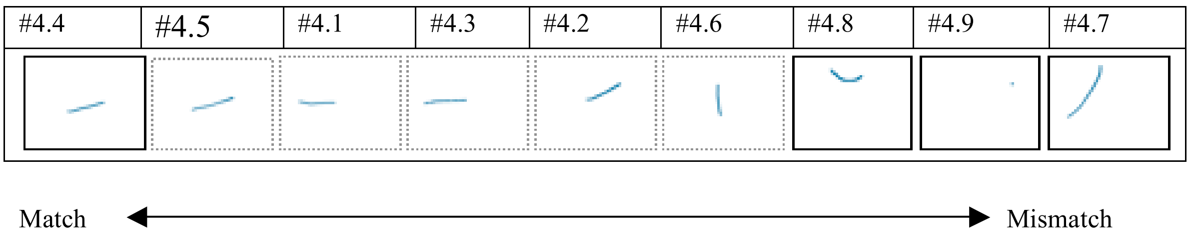


Figure 32. Drawing #4.4 is most consistently rated as the best match with the trumpet sound.

It seems then that the criteria used have something to do with the non-changing aspect of the trumpet sound, i.e. that the straight line visualises that pitch and timbre do not change. It might also be possible to interpret the lines according to imagery of the way the tone is produced, i.e. that a continuous force is used to bring out the trumpet tone, and the even, sustained movement with which an unbroken line is drawn resembles this sound-producing quality.

The ratings of the drawings corresponding to the piano tone show a slightly lower level of agreement. In this example some of the raters also prefer a straight unbroken line for match. The non-changing feature of pitch and timbre is visualised with a straight line, such as in drawings #3.1 and #3.8. The attack/decay envelope is not present in these drawings.

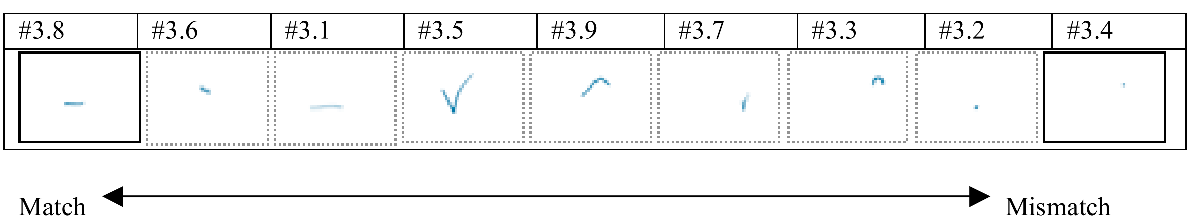


Figure 33. Drawing #3.8 is consistently rated as the best match for the piano sound.

There is a greater diversity in the drawings in this example compared to the trumpet sound example. This may reflect that some of the sound-tracers have attempted to visualise aspects of sound-production, i.e. the impulsive manner in which a piano tone is played. For example, the dots in drawings #3.2 and #3.4 have been made by pressing the pen on the tablet at one point, thus resembling the way a key is pressed on a keyboard. Similarly, drawings #3.5 and #3.7 may be understood as reflecting sound-production, but in a slightly different manner. The V-shapes seems to have been made with a quick downwards movement which resembles the first impulse or attack, followed by an upwards lift that illustrate the resonance phase of the tone. Regarding drawing #3.5 the judgements are spread since eleven (of twenty) judge it to have ‘good correspondence’ whereas seven raters consider the drawing to match the sound poorly. This disagreement may be interpreted in two ways: either the continuous feature of the tone is considered to be a more salient feature of the dynamics of sound-production, or the sound-production aspects are not properly expressed in the restricted pen and tablet format.

Sound example 11 is a very short speech sound resembling a “meh”. Considering the short duration, there is no prominent change in timbre and pitch, and the sound is produced impulsively. The corresponding drawings are quite similar to each other;, except for drawings #11.4 and #11.6 which both have a curvy shape, they are all horizontal lines, most of them very short, almost like dots.

There is very little consensus among the raters; the bar chart shows that the ratings are distributed across all categories of judgment for most of the drawings, e.g. drawings #11.3 and #11.6 are either judged as match or mismatch.

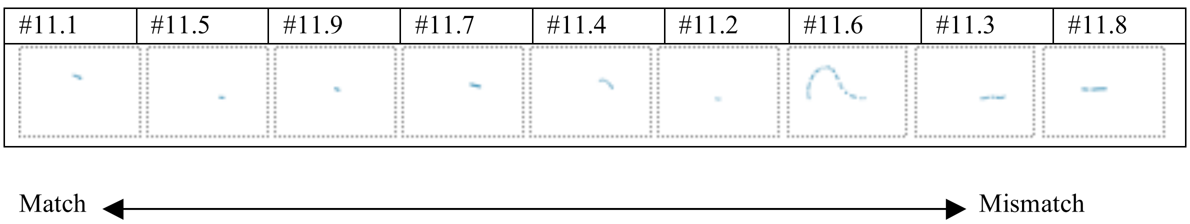


Figure 34. None of the drawings in example 11 are consistently judged as match/mismatch for the speech sound “meh”.

It is difficult to know how to interpret this uncertainty. One possible interpretation is that the drawings are too similar so that it is difficult to distinguish clearly between them[62]. Despite the observed disagreement, it seems that the duration at least may have played a role and led some of the raters to assign drawings #11.3 and #11.8 to the low correspondence category, thus judging match according to the amodal attribute of duration.

In example 16 the sonic element is the sound of a struck cymbal, typically characterised as an attack with a long decay, and with a total duration of 4.5 seconds. This means that the mode of sound-production is impulsive and the pitch is undefined. The change that characterises the sound may be described as a dynamical change (attack followed by a decay); however, one may also characterise it as a timbral change, i.e. from the ‘crash’ of the attack to a more thinned out timbral feature towards the end.

The drawings show considerable variation. We might speculate that the sound-tracers have had difficulties in visualising the sound, and that they have employed different

[62] One of the participants in the initial sound-tracing study noted that since this is a speech sound it is associated with feeling/emotion, so that the sound is too vivid to be captured within the simple drawing format

strategies. Again the ratings indicate a relatively low degree of agreement, although a little higher than in the speech/sound example.

The highest agreement on match is achieved in drawing #16.8. It seems that the sound-tracer has tried to bring out the attack/decay envelope of the sound in a similar way to what was done with the piano tone, i.e. a V-shape is combined with a tail resembling the downward movement that results in the attack of the tone and the following decay.

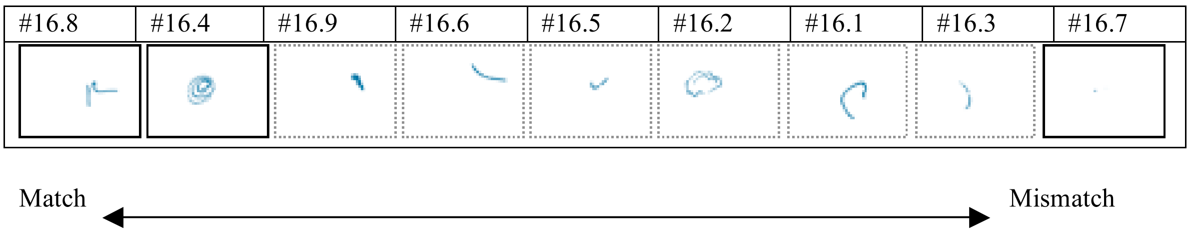


Figure 35. Drawings #16.8 and #16.4 are consistently judged as best match for the struck cymbal sound.

Regarding drawing #16.4, I assume that correspondence is perceived on the basis of timbral characteristics, i.e. that the light, scribbled quality of the drawing resembles the undefined porous and scratchy surface of the cymbal sound just after the point of attack.

To sum up this group of examples, the analyses suggest that raters prefer straight, unbroken lines to match the non-changing feature of for example the trumpet and piano sounds, but that motor imagery of sound production may also interfere with this criterion. The highest level of agreement is observed in the trumpet example, presumably because the *pitch contour vs straight line* correspondence converge with the *straight line vs continuous air pressure* match. The lowest level of agreement is found in the speech sound (“meh”). As previously suggested, the drawings may have been too similar, which presumably has led to random guessing, although in fact most of the raters spent quite a long time on this particular example. One of them commented that the speech sound appears richer than the instrumental sounds because it offers emotional content. Thus, the drawings were too sparse to express this feature.

Finally, the cymbal example demonstrates a different kind of correspondence, as it seems that the raters have recognised a match between features of the sound related to texture/timbre, i.e. the porous timbral feature and the light scribbling in two of the drawings.

6.4.2. Group 2. Pitch contours

For the next group I have chosen six examples that have a comparatively clear pitch contour. I assume this to be a salient feature of the sounds. I will take a closer look at how this feature interacts with and is affected by other features.

The first three examples are characterised by a descending pitch contour; i.e. glissandi on an electric guitar (example 5)[63], on a double bass (example 14)[64], and with a voice (example 13)[65]. The visual responses shown in figure 36 are in their general curvy lines and large extensions markedly different from the drawings discussed in group 1.

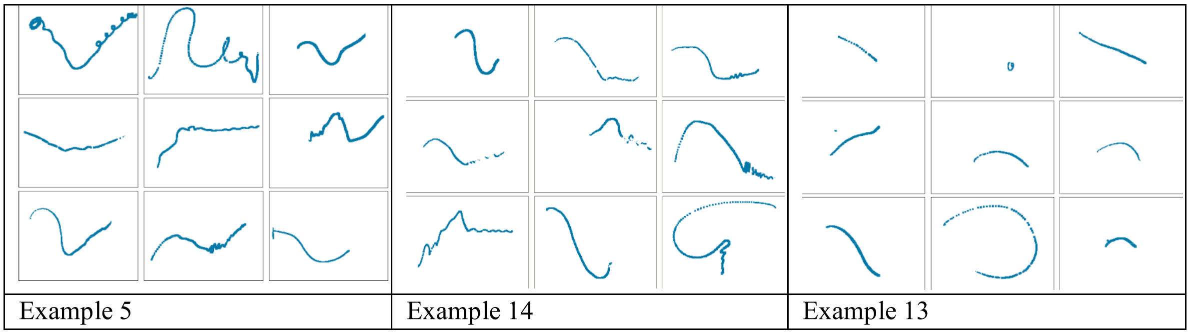


Figure 36. Drawn responses in examples 5, 14 and 13.

The sound used in example 5 is a drawn-out tone from an electric guitar that starts in a high register, followed by a descending continuous glissando, and when it reaches the bottom pitch in a lower middle register then ends with a slightly ascending glissando. The tone is played with varying degrees of distortion, with a little at the beginning and increasing considerably towards the end. Thus we have a change in timbre from a light, quality to what one might call a dirtier, rougher timbre towards the end (see spectrogram in figure 37).

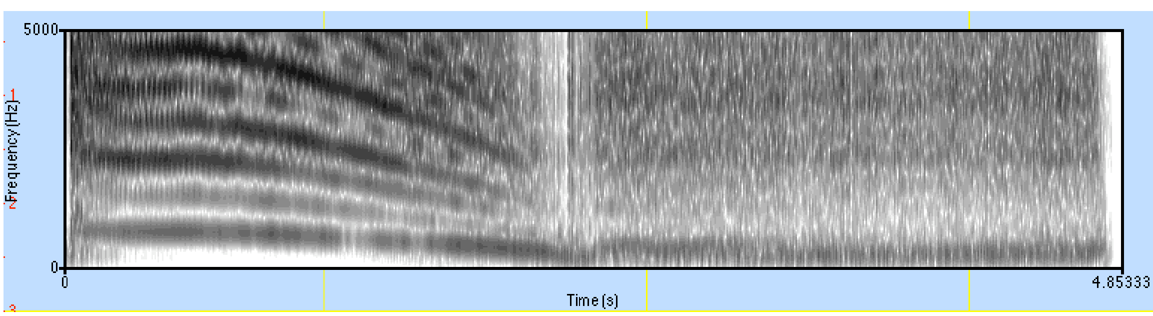


Figure 37. Spectrogram of the electric guitar sound used in example 5.

As a whole the tone appears to be performed through a continuous, sustained movement. However, the first part of the segment may be heard to have an impulsive quality; the tone is initiated with a clear attack, and since the following descending glissando has the quality of something falling freely downwards, the attack and the glissando together produce the impression of a ballistic-type event.

---

[63] From general sample bank
[64] Delalande, F., Formosa, M., Frémiot, M., Gobin, P., Malbosc, P., Mandelbrojt, J., et al. (1996). Les Unités Sémiotiques Temporelles: Éléments Nouveaux D'analyse Musicale [CD]. Marseille.
[65] From Pierre Schaeffer's *Solfege de l'Objet de Sonore*

The sound segment in example 14 is a compound of three main elements; a single triangle tone, a descending glissando on the double bass and a roll on a single tam-tam drum. It is initiated with simultaneous onsets on the triangle and double bass, followed by a glissando from a high register to a middle register. When the double bass reaches the lowest tone, the light roll of the drum takes over, attached as a kind of tail. Regarding production, this means that we have an impulse in the first attack, a sustained dynamic in the glissando, and an iterative aspect in the final drum roll. In terms of sound features, we have a change from high to low register, and a change in timbre from smooth to slightly rougher towards the end.

Example 13 is a speech sound, i.e. the sound of a female voice uttering an "ohh" sound with a falling pitch contour. There is no change in timbre. On the one hand it may be heard as being produced with a sustained dynamic, however, there is also an impulsive feature in the way the segment is performed; it is initiated with momentary force and the following sound appears to flow on top of this energy, suggesting that very little additional force is applied.

As noted above, the sounds in these three examples are similar in their descending pitch contour, but they differ in complexity. Example 13 has a 'clean' non-changing timbre, whereas the two others change in terms of timbre and in the introduction of additional timbral components. This is reflected in the drawn responses: the drawings in example 13 are made smoothly, which contrasts the rougher elements in the other examples.

In example 5, most of the drawings have traces of a U-shape, most apparent in drawing #5.1. There is also a tendency that these U's have a prefix in the shape of a short curve. Many of the drawings also alternate between an even line and a rougher, wavy tracing.

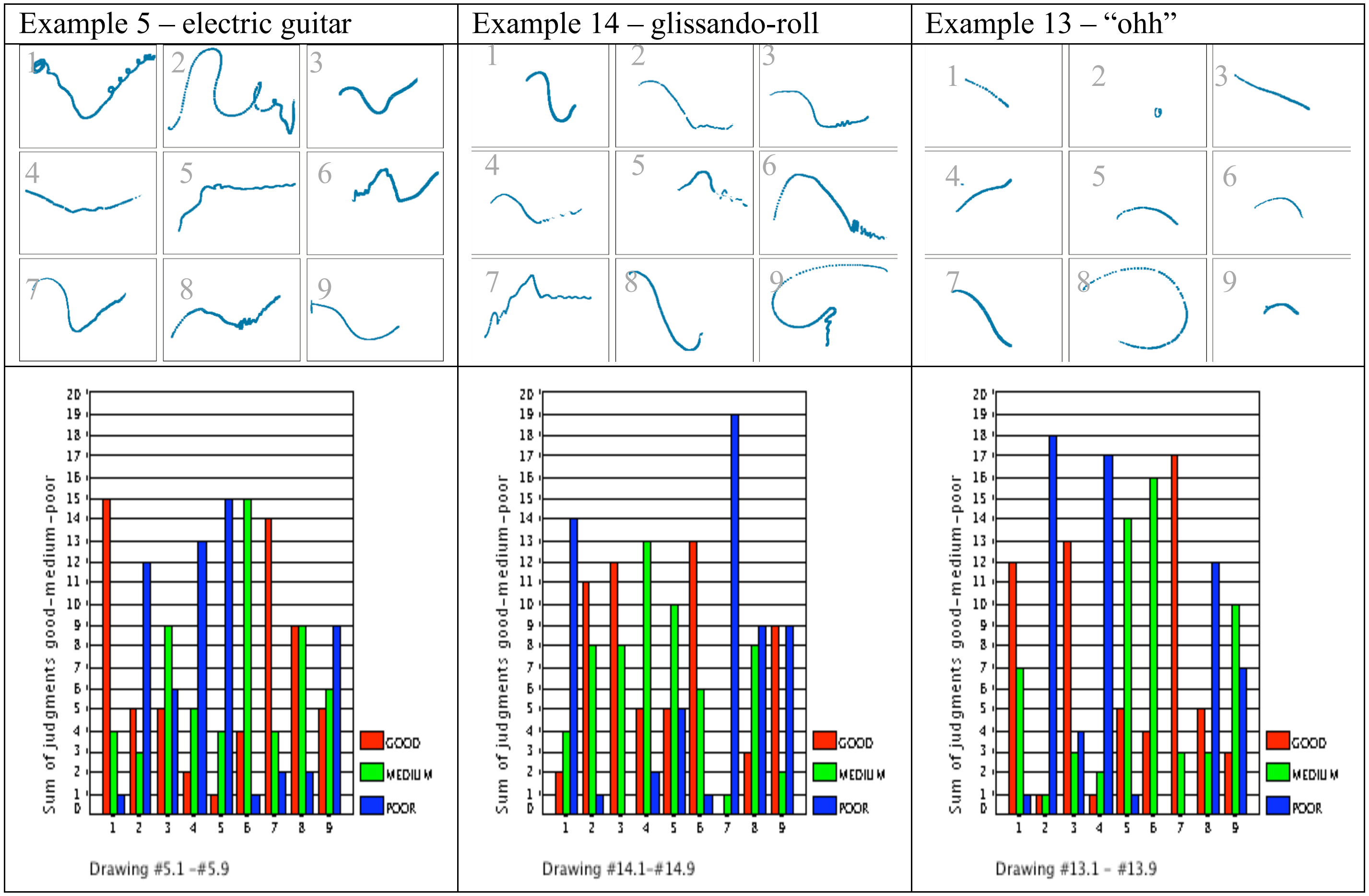


Table 12. The table shows the drawn responses to the sounds in examples 5, 14 and 13 together with bar charts summarising the ratings for each drawing.

As shown in the bar chart on the previous page, and in the series of drawings in figure 38, there is a relatively high degree of agreement in this example. It seems that the simultaneous changes in pitch and timbre are the main criteria for judgments of match, i.e. drawings #5.1 and #5.7 that combine these features are judged to have good correspondence.

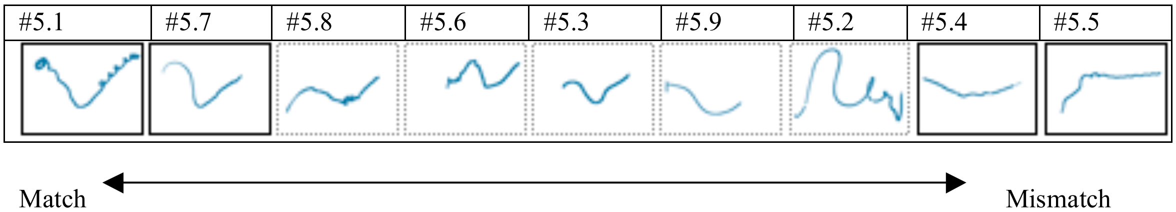


Figure 38. The U-shapes in drawings #5.1 and #5.7 are consistently judged to be best match for the electric guitar sound.

The two matching drawings both have a U-shape related to the continuous changes in pitch (high-low-higher) and a change from a smooth to a rough line corresponding to timbral changes. Drawing #5.1 even attempts to visualise the attack in the sound, but it seems that the judgments are mainly made on the basis of changes in pitch and timbre rather than production aspects. There is also good agreement that drawings #5.2, #5.4 and #5.5 match the sound poorly; #5.2 and #5.5 have contours that are difficult to relate to the pitch contour, whereas #5.4 appears to be too flat/levelled.

In example 14, except for drawings #14.7 and #14.9, all drawings have a direction which starts from a high position on the tablet, then goes downwards and to the right and ends in a low position. Additionally, they have a curvy quality. They are also characterised by the way they start with a horizontal line before they start their way downwards and then level out towards the end.

There is comparatively good agreement that drawings #14.3, #14.2 and #14.6 correspond well to the sound. They all have a characteristic arch-shape that seems to match the descending glissando, and they have a tail made up of a straight horizontal, uneven line, probably meant to illustrate the drum roll. Drawing #14.4 is quite similar but the contour is flatter, and it seems that the raters preferred a steeper fall as a better match for the glissando. Drawings #14.1 and #14.8 both have this steepness but they lack the rough tail, and this may be the reason why the judgments differ.

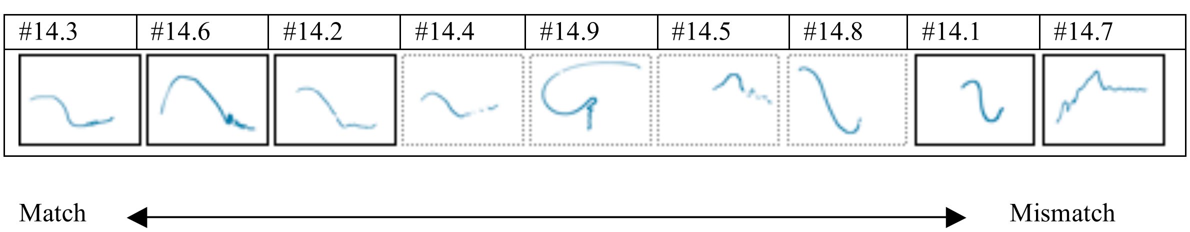


Figure 39. Drawings # 14.3, #14.6 and # 14.4 are consistently judged as the best match for the glissando sound.

There is full agreement that drawing #14.7 matches poorly, most likely because of the direction of the contour. The opinions about drawing #14.9 diverge; it appears to be confusing in terms of direction, though it combines the features of a curvy, smooth, even line with a rough tail.

The drawings of example 13 are generally characterised by lines of medium length that go in different directions on the tablet, and that also differ by being straight or lightly curved. All drawings seem to be drawn with an even, unbroken line. Two drawings stand out; drawing #13.8 almost completes a full circle, and drawing #13.2 is a small dot on the tablet.

The raters agree that drawings #13.1, #13.3 and #13.7 match the sound best, i.e. lines that are comparatively straight, and that are made with a movement that starts in an upper-left position and moves diagonally down to the right. Drawing #7 achieves the highest rating for match, possibly because it suggests a curved quality, which is also preferred in the other examples with a glissando (see example 14). Drawings #13.2, #13.4 and #13.8 are judged as mismatch, which points in the direction that pitch contour strongly affects perceptions of correspondence in this kind of material.

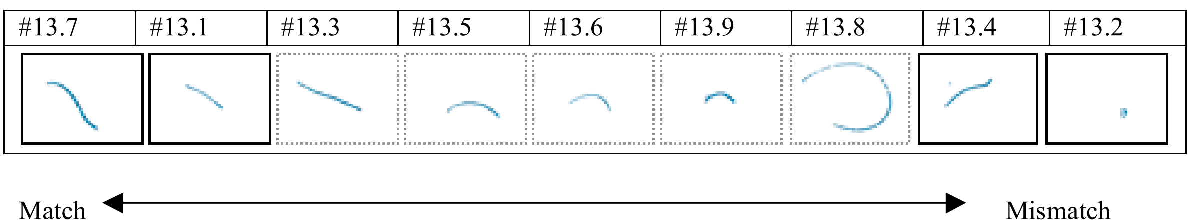


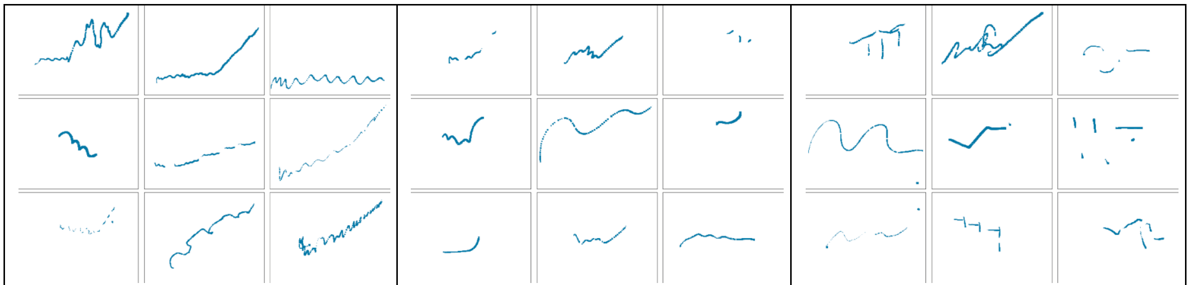
Figure 40. Drawings #13.7 and #13.3 are consistently judged to be best match for the descending sound.

In sum, the three examples (5, 14 and 13) in this group point in the direction that pitch contour and the way it is reflected in the trajectory, i.e. the kinematics of a movement, strongly affect the perception of correspondence in this kind of material. Thus, the raters prefer a descending, diagonal trajectory (from left to right). Additionally, it seems that the raters are sensitive to timbral changes (the elements of distortion in the electric guitar, example 5 visualised with a rougher line), as well as the aspects of sound-production introduced by the drum roll in example 14, i.e. the iterative sound-production is reflected in quick, scratching movements with the pen.

The next three examples are also characterised by their apparent pitch contours. The first two are musical phrases; example 12 is a short phrase played on the piano from Scriabin's *Sonata no. 5 op. 53* (the opening)[66], and example 15 is a phrase from Stravinsky's *Concertino for 12 Instruments*[67]. The third example is a speech phrase "il est

[66] From the recording: Austbø, H. (1989). Scriabin: Piano sonatas vol. 1. [Audio-CD] Simax PSC 1055.
[67] The excerpt is performed by The Danish Concert Orchestra.

déjà parti?”[68] performed by a female. Thus, the examples are characterised by successive tones/sonic elements that are grouped together into one phrase or chunk, the Scriabin and the voice examples with an ascending pitch contour and the Stravinsky example with a descending pitch contour. The corresponding drawn responses are shown in figure 41:



Example 12	Example 17	Example 15

Figure 41. The drawn responses for examples 12, 17 and 15.

The sound component in example 12 has duration of 5-6 seconds. It starts with a deep, rumbling passage played in the lower register of the piano, as though it is building up energy, and then moves upwards stepwise in the register. The density of tone onsets is high; figures played by both hands are interwoven, thus creating a dense texture. There is also a rhythmical drive towards the goal-point in the higher register, this sense of pushing forward emerging from variations in emphasis and articulation. The overall impression is a ‘mass’ of sounds glued together that move together upwards in pitch. However, there are also single tones that are more distinctly articulated, so that they stand out from the texture.

Regarding the speech phrase in example 17, it consists of six syllables that are heard as successive onsets. The timbre of the voice is smooth and non-changing, and the pitch contour appears to be evenly ascending. Furthermore, there is a change in onset density in time from the beginning to the end of the segment: the first five syllables are performed quickly (‘il-est-dé-jà-par’), whereas the last syllable (‘ti’) is more stretched in time.

The sound segment in example 15 is characterised by the way it consists of contrasting modes of production. A motif consisting of a long, sustained tone (high pitched flute) and a pizzicato tone (middle-register strings) are played three times, so that a pattern of long-short-long-short-long-short emerges. The phrase has an overall descending pitch contour. Furthermore, it differs from the others, as it seems that the musical texture is less concentrated.

All the drawings in example 12, except for #12.3 and #12.4, describe an upwards diagonal, from left to right. Another general feature is that they are drawn with a wavy, uneven line. The ratings show a high degree of agreement (see summarising bar charts in table 13).

[68] From Pierre Schaeffer’s *Solfège de l`Objet Sonore*

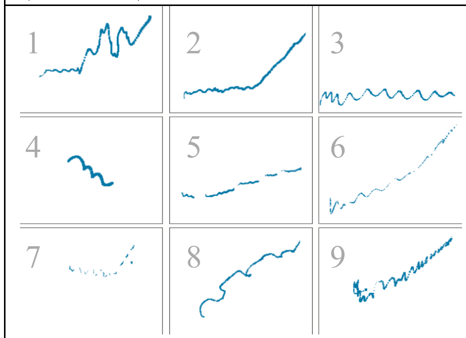
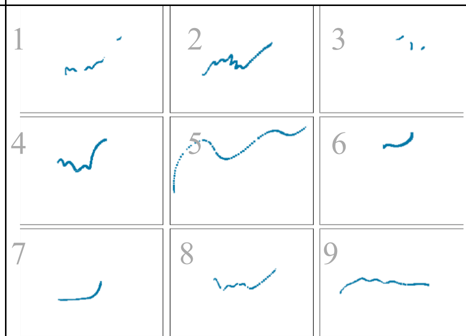
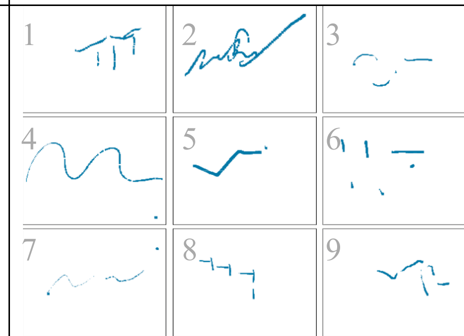
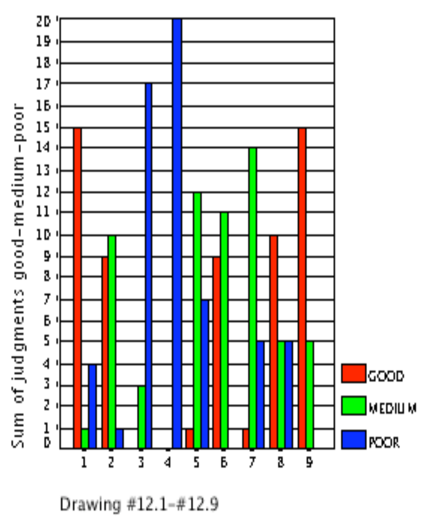
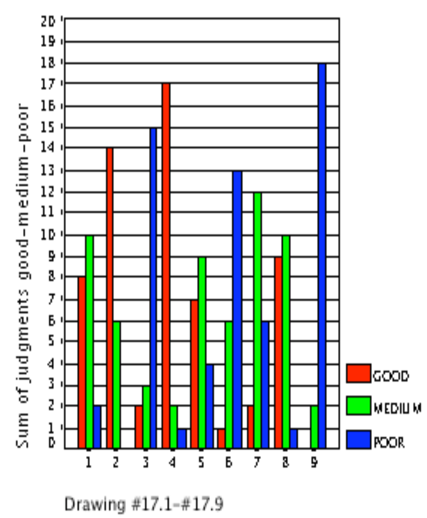
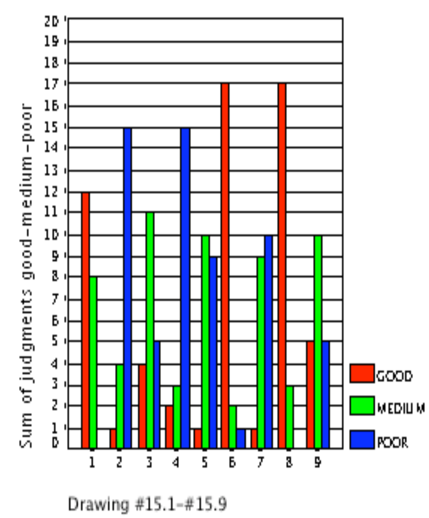
<table>
<tr><th>Example 12 – piano phrase (Scriabin)</th><th>Example 17 – ”il est déjà parti”</th><th>Example 15 – Stravinsky</th></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

Table 13. The table shows the drawn responses to the sounds in examples 12, 17 and 15 together with bar charts summarising the ratings for each drawing.

Drawings #12.1 and #12.9 are judged to match; both combine a rising contour (left to right) with a rough, jagged line. Presumably the contour corresponds to the global change in pitch from low to high, whereas the uneven line attempts to visualise the onset density and maybe also the occasional, distinctly articulated tones. Drawings #12.3 and #12.4 are generally judged as mismatch, probably because the overall direction of the line does not correspond to the direction of the pitch contour.

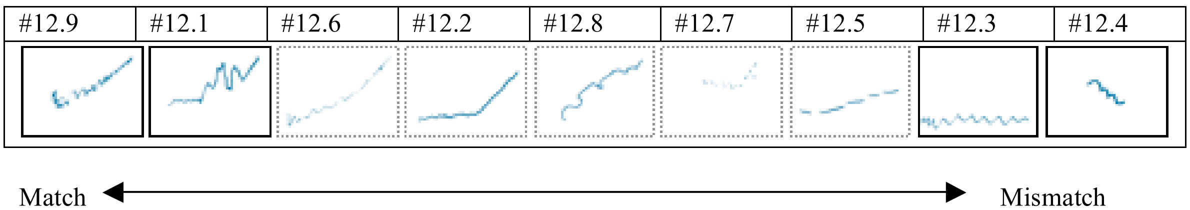


Figure 42. Drawings #12.9 and #12.1 are consistently rated as the best match for the ascending piano contour.

Similarly, most of the drawings in example 17 suggest a rising contour, seen from left to right. Only drawings #17.3 and #17.9 do not have this feature, and they are both relatively consistently rated as poor match. The three drawings (#17.4, #17.2 and #17.8) that are agreed to correspond well to the sound seem to combine three features: first, they have an overall rising contour; second, they have an abrupt upswing towards the end (to the right); and third, they are characterised by a rougher line at the beginning (to the left) that changes to a smoother quality. This rougher line, performed with rapid hand movements probably reflecting the high density of onsets in the sound, and the smoother line, the upswing towards the end, most likely are meant to visualise the concluding, relatively long 'iih' sound

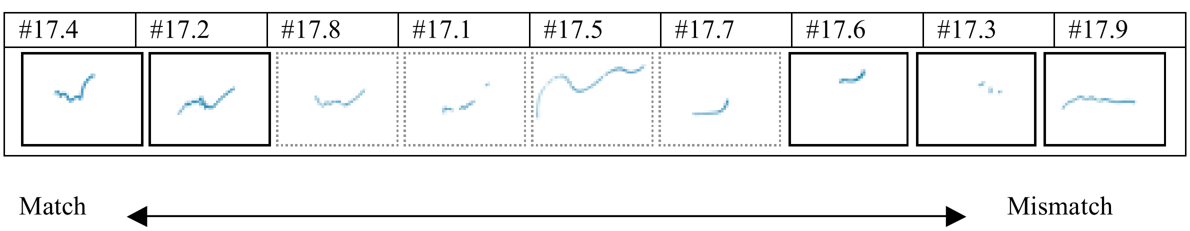


Figure 43. Drawings #17.4 and #17.2 are consistently judged as the best match for the "il est déjà parti" phrase.

In example 15 there is marked agreement on match-mismatch. Drawings #15.6, #15.8 and #15.1 are judged as match, and drawings #15.2 and #15.4 are considered to match the sound poorly.

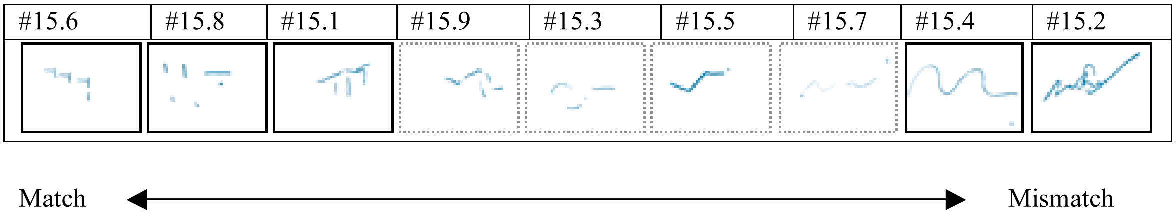


Figure 44. Drawings #15.6, #15.8 and #15.1 are consistently judged the best match.

Thus, it seems that the raters prefer the drawings that illustrate the contrast between long (woodwind) and short (pizzicato) tones, and the way these are grouped. Alternatively, the best-match drawings may be interpreted within a production perspective, i.e. the long lines reflect motor imagery of the sustained air pressure that produces a long woodwind tone, whereas the dots in #15.6 represent the abrupt movement employed in plucking a string. In drawings #15.8 and #15.1 it appears that the pizzicato feature is visualised with a vertical line, thus alluding to the vertical movement used to pluck a string. Furthermore, the lines and dots are more spread vertically than in the other examples, and I would suggest that this feature attempts to bring out a textural/pitch space aspect, i.e. that the texture is less concentrated in this example. Finally, the overall pitch contour is vaguely suggested in drawings #15.6 and #15.8, whereas the 'mismatch' drawings are drawn with a contour that seems to contradict this global feature.

To sum up these two first groups of examples, those characterised by non-changing pitch and those by an apparently ascending or descending pitch contour, the analyses point in the direction that the *kinematics* of sounds (i.e. understood as pitch contours, see chapter 4) and drawings (i.e. understood in terms of trajectory) strongly affect judgment of correspondence. Furthermore, the match judgment is strengthened when this feature converges with aspects of sound-production. This is seen in the Scriabin example where the best match drawings combine an upward trajectory with a jagged line, this drawing-feature presumably alluding to the way some of the piano tones are being produced with distinctly articulated finger-movements.

In the Stravinsky example, however, it seems that the global pitch contour plays a less prominent role, and that other aspects such as grouping, sound-production/duration contrast, and texture are more important for judging correspondence. In the next group of examples I shall discuss three examples that are characterised by their grouping features.

6.4.3. Group 3. Patterns of successive sounds

The sound segments used in examples 2, 6 and 9 are all recordings of birds' twittering[69], and are characterised by being groupings of individual sonic elements. Examples 2 and 9 consist of distinctly repeated and identical short sounds, (in example 2, five successive

[69] All examples are from Pierre Schaeffer's *Solfège de l`Objet Sonore*

short sounds, and in example 9, three short sounds). Together these single elements form a sonic pattern. There is no global change in pitch or timbre in the two patterns, but there is a fall in pitch in each of the single sounds. The two examples differ as the sounds in example 2 are performed with a higher onset density in time. Because of this the fall in pitch in the single sounds is much more stretched out and marked in example 9.

Example 6 appears to be a more complex pattern. The whistling sounds are not as distinctly articulated or separated byn time compared to the previous examples so that the single sounds merge into a more coherent stream. Within this stream there are changes in pitch up and down which means that the sounds occupy a larger pitch space than in examples 2 and 9. The overall pitch contour also makes a slight curve upwards at the end of the pattern. Timbre and dynamics are not prominent features.

The observed difference in sounds, the separate vs the merged, is reflected in the drawings as shown in figure 45 which also includes spectral analyses of the sounds:

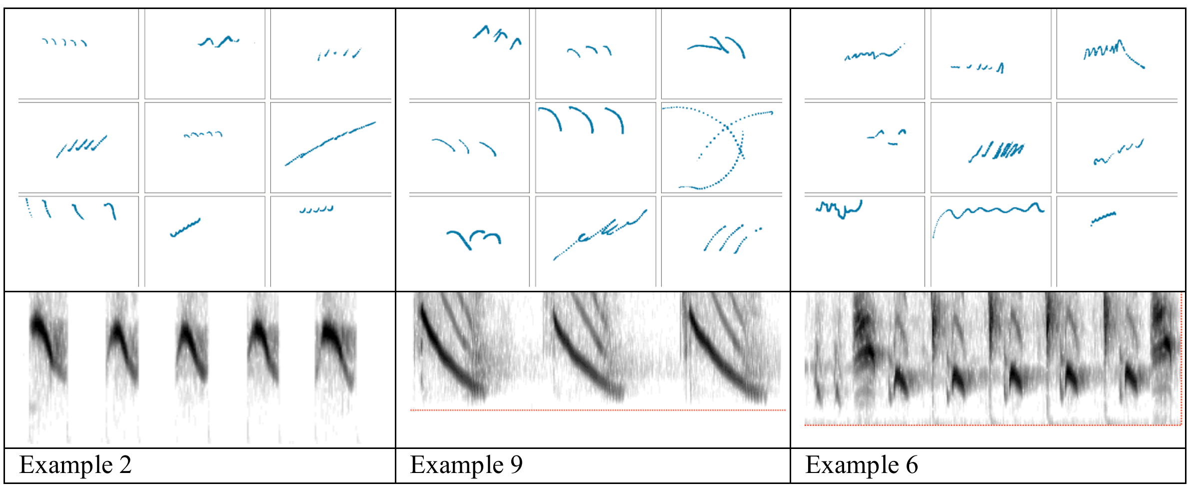


Figure 45. Drawn responses in examples 2, 9 and 6, and the spectrogram analyses of the corresponding sounds.

Most of the drawings in example 2 consist of repeated short, lightly curved lines. The exceptions are drawings #2.6 and #2.8, which both have an unbroken, diagonally directed line. The ratings of these examples are shown in table 14.

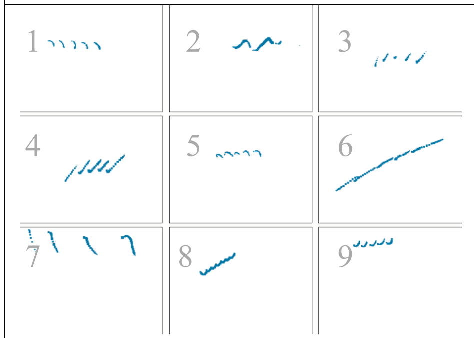
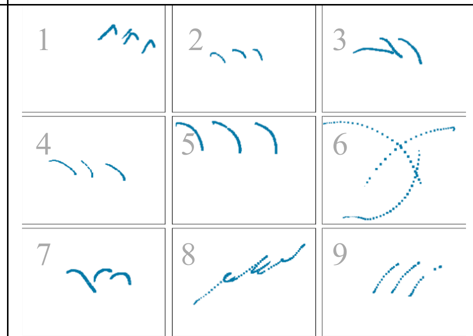
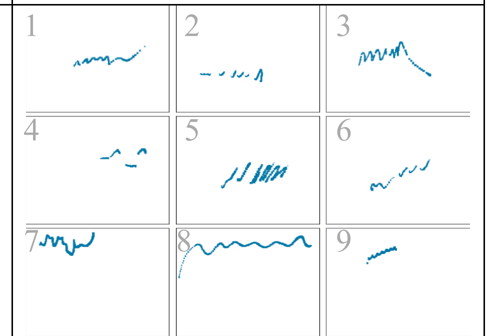
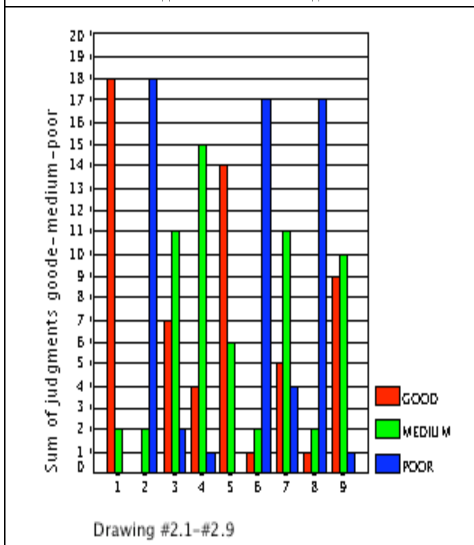
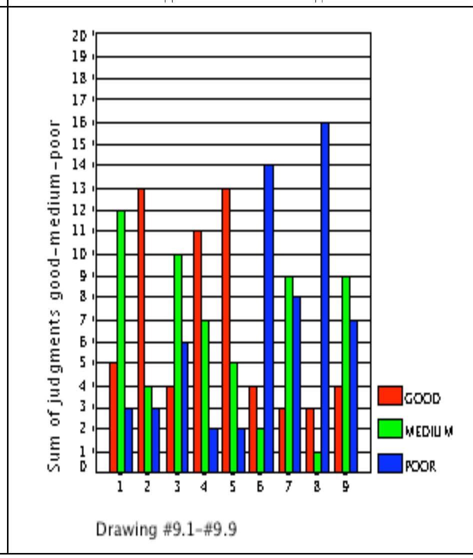
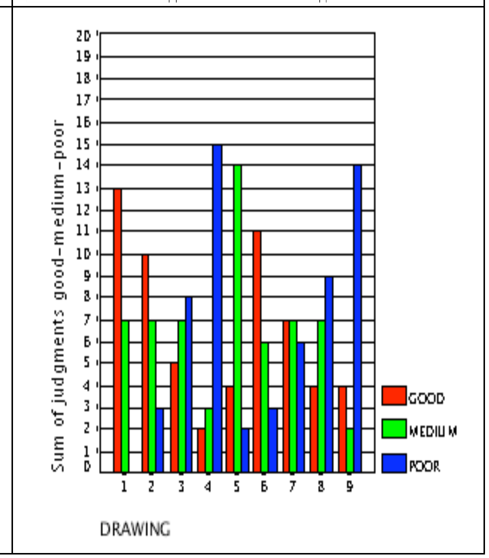
Example 2 – bird	Example 9 – bird	Example 6 - bird
		
		

Table 14. The table shows the drawn responses to the sounds in examples 2, 9 and 6 together with bar charts summarising the ratings for each drawing.

There is marked agreement between the raters. Drawings #2.1 and #2.5, and #2.9 are rated to match. It seems that the judgment is based on the number of strokes (thus corresponding to the number of impulses in the sound pattern), as well as the downwards curve in each stroke (matching the slight fall in pitch within each sound impulse).

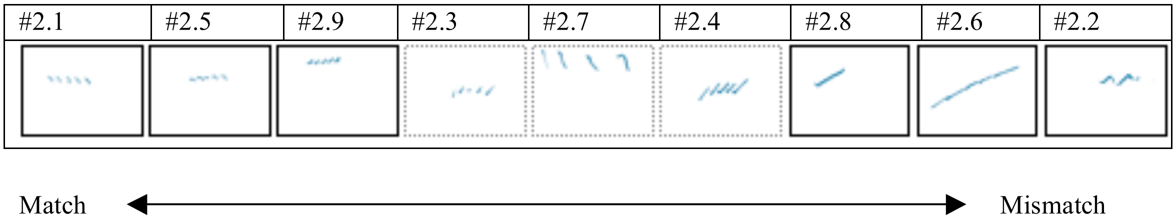


Figure 46. Drawings #2.1, #2.5 and #2.9 are consistently judged best match for the first bird sound pattern.

Also drawings #2.3 and #2.4 preserve the correct number of strokes, but the direction of each curved stroke appears to be wrong in these drawings. Additionally, one may assume that the overall direction of the visual patterns is connected to the overall non-changing pitch in the sonic pattern as a whole. In line with this, drawings #2.8 and #2.6 are agreed to match the sound poorly.

Similar to example 2, the drawings tend to correspond amodally to the sound; i.e. with respect to the number of strokes. Three drawings, #9.5, #9.4 and #9.2, are agreed to match the sound best. They are all characterised by a pattern of three downwardly curved lines. Thus, it seems that judging correspondence is based on number and kinematics (i.e. a descending pitch contour corresponds to a diagonally descending trajectory, seen from left to right).

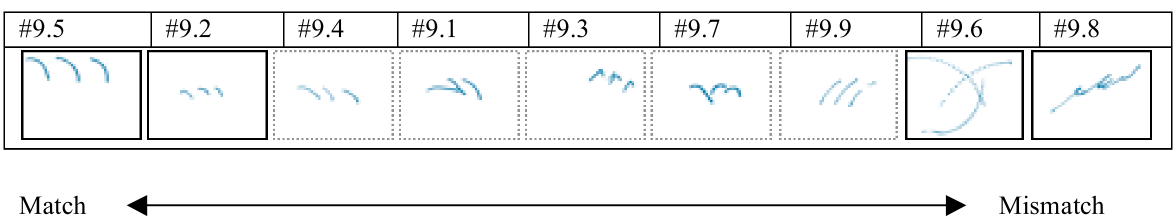


Figure 47. Drawings #9.5 and #9.2 are consistently judged as the best match for the second bird sound pattern.

All drawings in example 6 appear to have been made with short, abrupt scratching movements. The line is in some cases even and in others broken up. There are also differences in the character of the short curves; some of them are rounded and smooth whereas others are more like sawtooth-patterns.

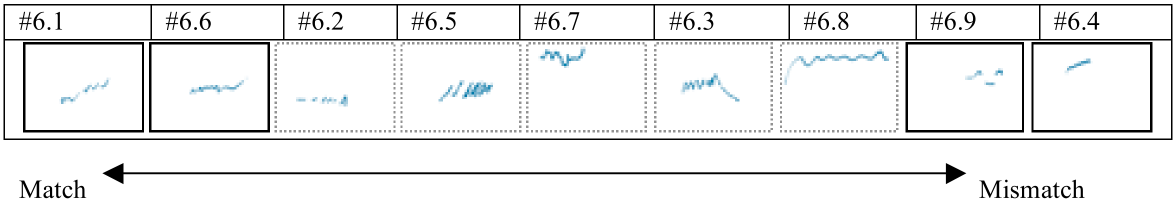


Figure 48. Drawings #6.1 and #6.6 are consistently judged as the best match for the third bird sound pattern.

Drawings #6.1 and #6.6 are judged as match. Both seem to visualise recurrent sound impulses with a jagged line, and in both cases the finishing pitch change is recognised in an upwardly directed contour (seen from left to right). Accordingly, drawing #6.3 seems to be judged in the direction of mismatch since it has a downwards tail at the end, which contradicts the pitch change.

The raters also agree that drawing #6.9 matches poorly. The drawing seems to be too short, and to be made with too subtle scribbling movements to match the duration and the pitch space of the sound. Finally, the judgments also point in the direction that the raters have been sensitive to the jaggedness of the lines, e.g. the raters may have considered the curves in drawing #6.8 to be too rounded, and the strokes in drawing #6.5 to be too sharply articulated.

The most prominent matching criterion for this group of examples is the amodal attribute of number, i.e. the number of strokes in the drawings corresponds to the number of sound impulses. Additionally, this feature is combined with the length of each stroke corresponding to the duration of sound impulses (duration also an amodal attribute), as well as the downward direction of the strokes, which matches the descending pitch contours. In general, there is a high level of agreement, and we may assume that the clear patterning that characterises the examples and the amodal criteria for correspondence afford movement unambiguously.

6.4.4. Group 4. Iterative sound-production

I have assigned two examples to a group that is characterised by an iterative mode of sound-production. One typical example of iterative sound-production is a drum roll, i.e. the sound evolves as a process of successive impulses/strokes with a high density in time. Technically, the high density is managed by keeping the wrist loose, thus taking advantage of a rebound effect. In other cases the iterative sound may be produced with one single movement, such as the sound made when a stick is pulled along a fence made of vertical boards.

The sound in example 1 simulates the sound of a mallet that is set in motion between two metal plates so that it hits the two plates at a high frequency[70]. This means that the

[70] Created in *Modalys* by Rolf Inge Godøy

mode of production is mainly iterative, so that a rolling feature is achieved. There are also other aspects involved; the sonic process is initiated with an impulse, and moreover, because a subtle swell in dynamics/loudness occurs just after this first impulse, we get the impression that energy is continuously supplied during this phase. The whole segment lasts for about five seconds, which means that we have a main impulse consisting of iterative impulses that last for a little more than half a second, followed by a long, freely resonating, 'natural' decay. The pitch is stable or undefined throughout, and similarly the timbre is non-changing.

In example 10 the sound is created by a rainstick[71], i.e. a digital version of the original acoustic African instrument. The sound resembles the sound of a very dense texture of numerous raindrops falling to the ground, but the sound is much drier than those produced by real raindrops. Thus the timbre has a non-changing light, airy, trickling feature. And similar to example 1, the pitch is undefined and there is no change in dynamics.

The drawn responses are shown together with rating bar charts in table 15:

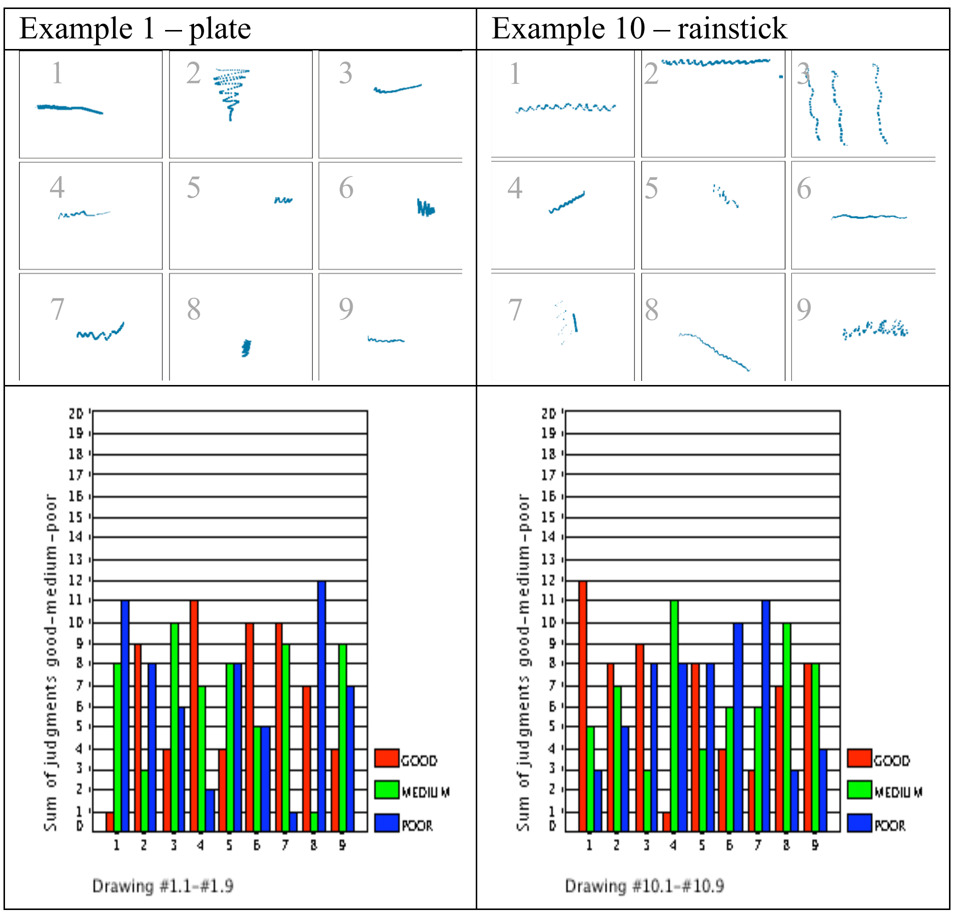


Table 15. The table shows the drawn responses for the sounds in examples 1 and 13 together with bar charts summarising the ratings for each drawing.

[71] From Max-tutorials/Cycling '74

An initial glance at the displays gives the impression that the iterative sound-production is reflected in quickly scribbled lines. This is apparent in example 1 in which all the drawings are to varying degrees made jaggedly. Furthermore in this example the drawings are, with two exceptions (#1.2 and #1.8), made with an overall horizontal movement.

Drawings #1.4 and #1.7 are judged as match by many of the raters, although it should be noted that the agreement is at a relatively low level. This seems to confirm that the raters have emphasised the iterative aspect of sound-production, i.e. the quick, small, up-down movements with which the jagged lines must have been made correspond to the repeated hits on the plate with which the sound has been made. They all illustrate the iterative aspect in different ways.

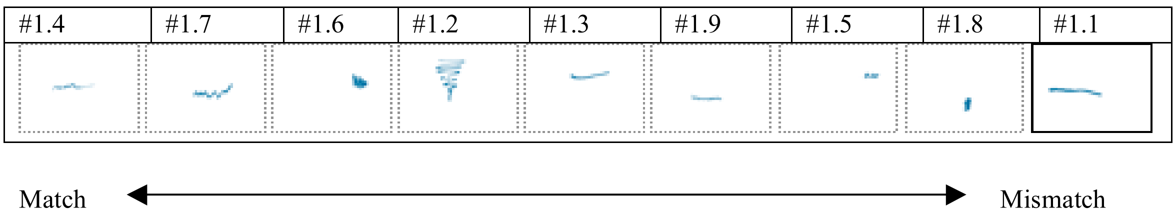


Figure 49. None of the drawings in example 1 are consistently judged as best match.

Drawings #1.2 and #1.4 also suggest the decay towards the end of the sound segment as they attach a straighter line as a tail after the rough movement. In line with the emphasis on the iterative aspect, the raters also agree that drawing #1.1 matches the sound poorly, probably because the line is drawn more evenly so that it does not correspond to the sound-production feature. The preference for a jagged line before an evenly drawn line can be shown better when the best-match drawing #1.4 is blown up and compared with the poor-match drawing #1.1 (figure 50):

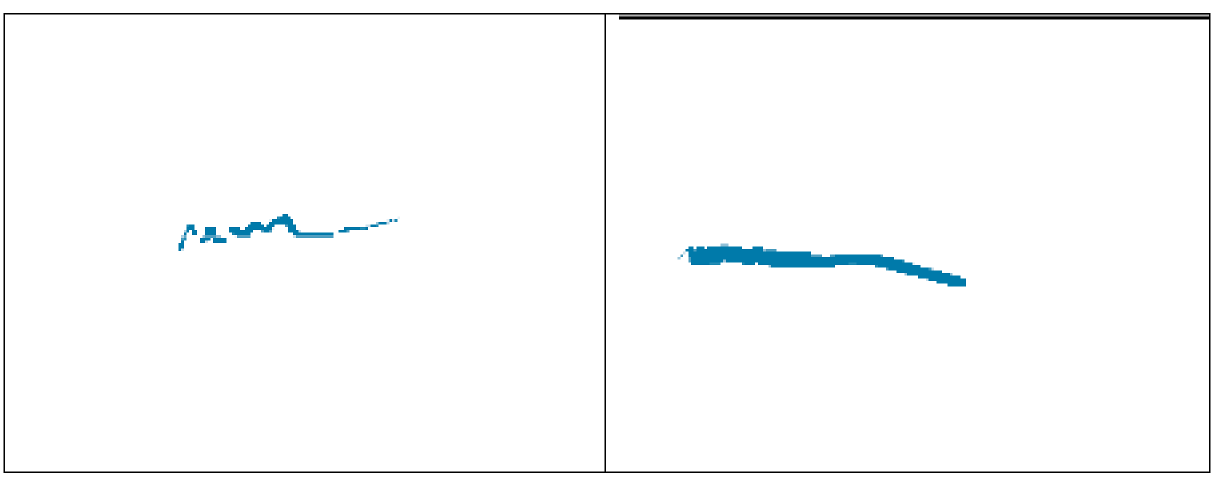


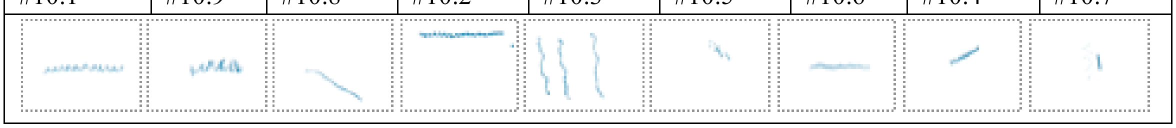
Figure 50. In example 1 where the sound is characterised by the iterative mode of sound-production the raters seem to prefer an unevenly drawn line for good correspondence compared to an evenly drawn line.

As noted above, there is generally a low degree of agreement. One explanation for this may be that there are competing features in sound-production, i.e. that the sound at the same time alludes to impulsive, continuous as well as iterative processes. Thus, it may initially have been difficult for the sound-tracers to visualise all these aspects convincingly, so that the succeeding ratings are based on chance rather than reliable choices of best/poor match.

In example 9, the drawings consist of lines with an overall non-curved contour which are drawn wavily and, flutteringly. In one of the drawings there are three such lines. Drawing #10.9 stands out as it seems to have been made by making numerous single dots with quick-handed movements within a rectangular area. A similar approach may have been used in drawing #10.5.

In this example we may also observe a considerable spread in the ratings. Drawing #10.1 is rated match but not with a relatively high level of agreement. It is characterised by a long, horizontal wavy line, and the perceived correspondence is probably based on the way the recurrent waves visualise the iterative aspect of the sound. On the other end of the scale, drawing #10,7 is agreed to match the sound poorly, probably because it lacks the iterative aspect. Similar to drawing #10.4, it appears also to be too short and 'heavy', so that it does not match the duration of the sound as well as the light and airy timbral character

#10.1	#10.9	#10.8	#10.2	#10.3	#10.5	#10.6	#10.4	#10.7



Match ◄──────────────► Mismatch

Figure 51. None of the drawings in example 10 are consistently judged as match/mismatch.

In sum, the most prominent criterion for correspondence in both examples seems to be the way the iterative mode of sound-production is reflected in successive articulations with the pen on the tablet. Second, in example 10 a further matching aspect may be found in the opposition between light and more heavily drawn lines, i.e. the light, dry timbre of the rainstick sound is recognised in the light quick manner of drawing in the best-match responses.

6.4.5. Group 5. Sounds with changes in dynamics and/or timbre (non-changing or undefined pitch)

In the final group I shall discuss examples 7, 8 and 18. Here, the sound elements resemble each other in the way they include dynamic and/or timbral changes, though with a non-changing or undefined pitch feature. The drawings are similar to each other in their relatively large contours, i.e. they occupy a large amount of space on the drawing tablet.

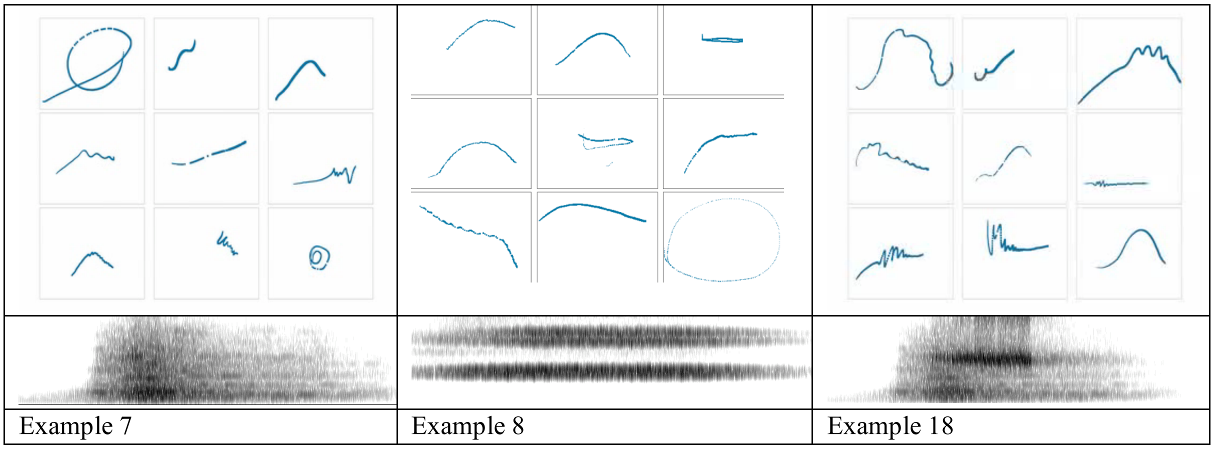


Figure 52. Drawn responses in examples 7, 8 and 18, and the spectrogram analyses of the corresponding sounds.

The sound in example 8 may be characterised as a light and airy one with a thin timbre that grows gradually in intensity and that also decreases gradually in intensity towards the end[72]. Thus it has a dynamic contour similar to the two other examples in this group, but in example 8 the contour is more stretched out and there is no apparent goal point. The pitch is stable throughout in the upper-middle register, and the timbre may also be described as having a firm core with airy, lighter, porous timbral elements surrounding it.

The sound in example 7 lasts for approximately 4 seconds. The pitch is difficult to define, and the timbre is dark and has a metallic character[73]. Initially, the sound is soft but swells to a climax approximately 1.5 seconds before it gradually fades away. This swell is the most prominent quality of the sound; it evolves as though something were pressing out through a narrow passage, i.e. it has a strained, bound, compressed quality at the beginning; and at the moment when it is released, the energy flows more freely, which means that it is heard as a momentary, abrupt increase in intensity. Then the intensity decreases more gradually. Initially, the timbre is dark, but towards the end it also gains some lighter elements. It also becomes slightly rougher during the decay phase.

The sound in example 18 is a variant of the previous[74]. It is based on the darkly swelling, metallic timbre, but a lighter timbral element is added on top of this basic tone. This element is introduced at the moment when the tension is released, and is characterised by an iterative sound-production feature, as well as a higher pitch and a rougher timbral surface. In example 8, six of the drawings are long-stretched, curved lines; some of them drawn firmly and evenly, others with a lighter, slightly rougher feature. Drawing #8.9 stands out as it describes a full circle. The two drawings #8.3 and #8.5 differ even more from the others, as they seem to be drawn with a straight, horizontal back-and-forth movement.

[72] From general sample bank
[73] From Pierre Schaeffer's *Solfège de l`Objet Sonore*
[74] From Pierre Schaeffer's *Solfège de l`Objet Sonore*

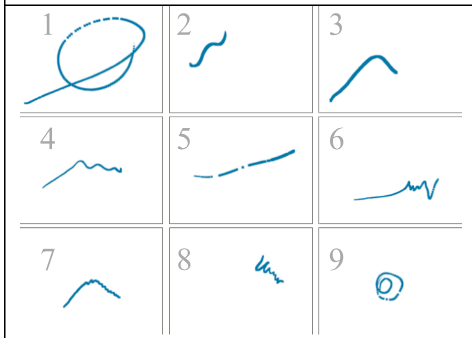
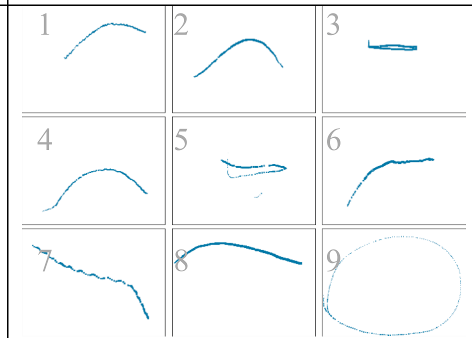
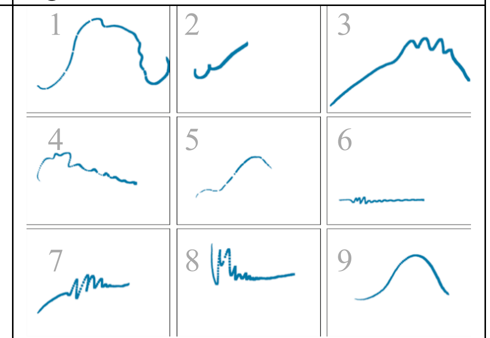
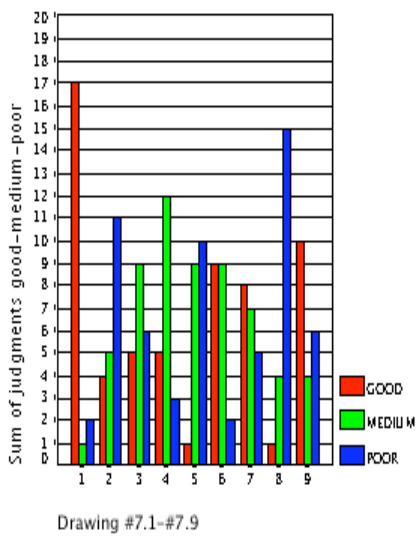
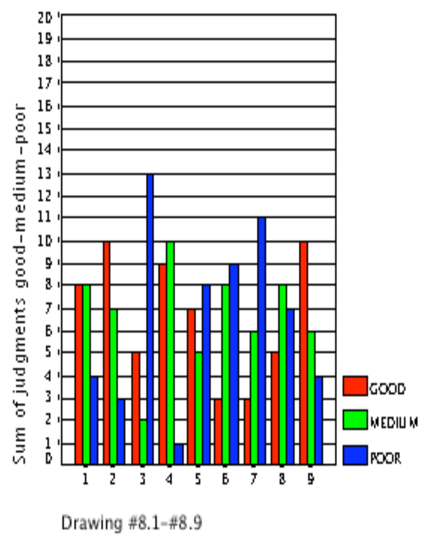
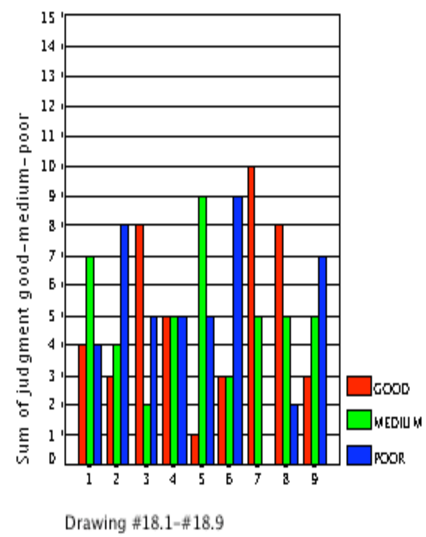
<table>
<tr><td>Example 7 – dark, metallic</td><td>Example 8 – light, airy</td><td>Example 18 – dark, metallic with lighter timbral element</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

Table 16. The table shows the drawn responses to the sounds in examples 7, 8 and 18 together with bar charts summarising the ratings for each drawing

All three examples in this group demonstrate a considerable spread in the raters' judgments. In the overview in figure 53, drawings #8.4 and #8.2 are ranked as match, but with a low level of agreement. Likewise, the mismatch drawing #8.3 is also rated non-consistently.

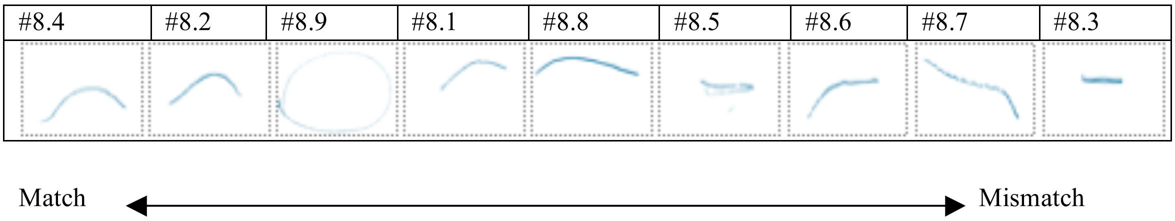


Figure 53. None of the drawings in example 8 are consistently judged as match/mismatch.

With respect to drawing #8.9, I would suggest that the full circle, drawn with a thin line, presumably made with a big, light and quick movement with the pen captures the airy, ethereal aspect of the timbre and that correspondence is based on this connection. Since the dynamic contour is not preserved in the visual contour, one might suggest that the sound/drawing relation is based on a metaphorical kind of correspondence.

One might have expected the long-stretched arched contours to have scored more consistently on match when bearing in mind that the shape of the changes in intensity appears to be a very prominent aspect of the sound. But there is a considerable spread in judgments of the drawings that have this contour (drawings #8.1, #8.2, #8.4 and #8.8). The disagreement applies also to drawings #8.1 and #8.4, which both attempt to combine the arch-contour with a light, rough line, thus attempting at the same time to visualise both dynamical and timbral features. It is difficult to interpret this disagreement; maybe the timbre and the sound as a whole are too rich to capture in this restricted format

Example 7 is characterised by relatively big differences between the drawings. Three drawings have a global ∩-contour, which is most apparent in drawings #7.3 and #7.7, and less apparent (flatter, more stretched out) in drawing #7.4. Two of the drawings may be described as diagonal contours (#7.2 and #7.5); two drawings are characterised by a sawtooth shape, brought out by a quick, scratching movement with the pen (#7.6 and #7.8); and finally, two drawings are made with a circular movement (#7.1 and #7.9). It is reasonable to believe that the ∩-contours are an attempt to capture the dynamic swell of the sound. Also, the big circle in drawing #7.1 may be interpreted as relating to the dynamics of the sound, i.e. if we imagine that the drawing has been produced with an initial slow movement and then followed by a big and quicker circular movement with the pen. Regarding the scratching feature that is found in some of the drawings, it seems that this attempts to visualise the subtle roughness of the timbre towards the end of the sound segment.

The highest score for match is obtained by drawing #7.1 probably because it convincingly visualises the dynamic swell. On the other end of the scale drawings #7.2 and #7.5 are judged as mismatch; they appear to visualise the sonic event rather incompletely. Drawing #7.8 is also rated low with respect to good match; it is made with a relatively big scratching movement that seems exaggerated compared to the fine nuance of the timbral roughness. Another explanation may be that the drawing is too small considering that the sound's dynamic swell would be associated with quite a large movement, i.e. the drawing does not capture the global quality of the sound.

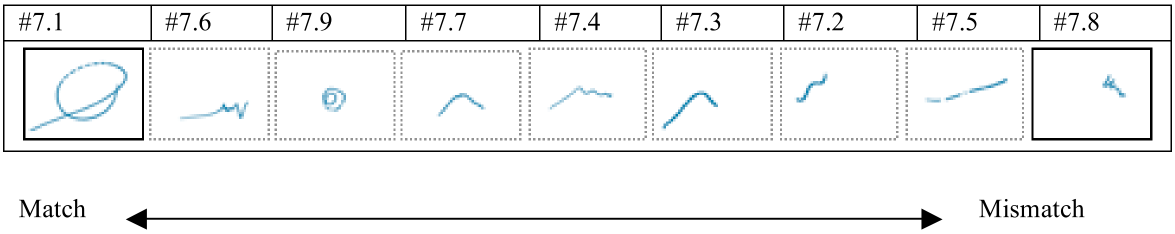


Figure 54. Drawing #7.1 is consistently judged as the best match for the dark, metallic sound.

As in the previous example, many of the drawings in example 18 have a global $\cap$-contour. It seems that the drawings, e.g. #18.7 and #18.3, that combine this shape with an element of jaggedness are judged as matching the sound[75]. Thus, we may assume that the contour of the drawings is perceived to match the dynamical contour of the sound. The mismatch rating of drawing #18.6, which is characterised by a flat contour, suggests that the arch shape is an important criterion of judgment. Furthermore, it seems that an unbroken line corresponds to the initial smooth timbral quality and that the more curvy elements correspond to the rough timbral feature introduced in the 'middle' of the sound segment.

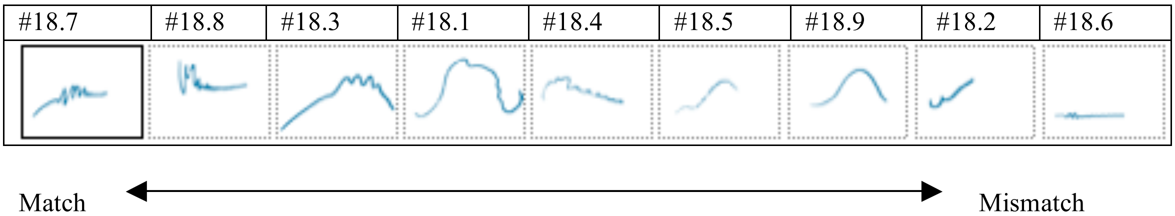


Figure 55. Drawing #18.7 is consistently judged as the best match for the sound characterised as dark, metallic with a lighter timbral element.

In sum, the examples in this group demonstrate a correspondence between arch-like shapes in the drawings and dynamic swells in the sounds. We may also observe that the raters respond to timbral aspects, e.g. *even, unbroken line vs smooth, firm timbre* and *light wavy manner vs airy, porous timbre.* There is generally a low level of agreement in the ratings. One possible interpretation is that the dynamic swells are not convincingly captured in the restricted two-dimensional format.

---

[75] As mentioned earlier in this chapter, the first five raters of this example had to be excluded for this example. Despite this weakness, I have chosen to include the example in the analysis since it is interesting qualitatively. This means that the analysis is based on the ratings of 15 raters; accordingly, the threshold level for relatively high level of agreement has been set to 7 instead of 10 for this example.

6.5. Conclusions and future directions

I have proposed the general hypothesis that sound affords movement on the basis of the way features change over time. In light of this, I have in the analysed sound-tracing material observed correspondences with regard to change in the following features:

- *Change of pitch.* The raters are sensitive to the directions and shapes of the drawings, e.g. demonstrated in the way they distinguish between rates of change (abrupt/gradual) in a curve. In example 5 they preferred the steeper curve, i.e. the drawing to the left:

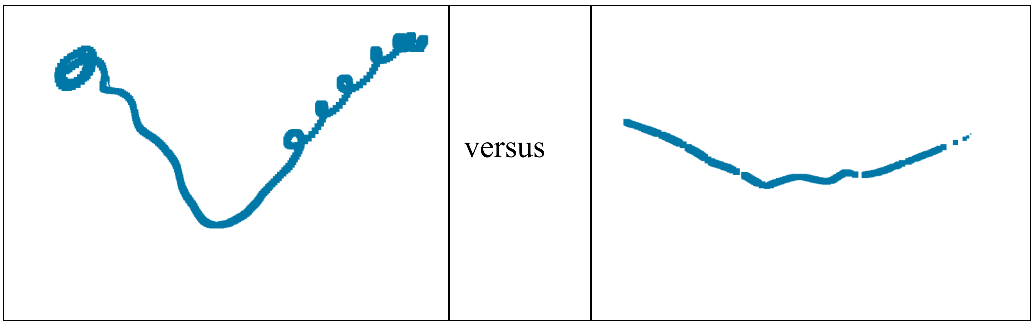


Figure 56. The raters were sensitive to the steepness of the curve when rating the drawings in example 5.

- *Changes of timbre* are reflected in the manner of drawing (even/broken/jagged/light/firm).
- *Changes in production* correspond for example to the length of the line (*straight line vs. sustained production*), and a jagged/rough line is matched with iterative production.
- *Changes in onset density of sound impulses* are reflected in the jaggedness of the drawings.

Additionally, the material exemplifies the way correspondences are perceived with respect to amodal attributes and patterning features. Based on the observed differences of agreement between the groups of examples, I would suggest that amodal transference and patterning similarities, together with pitch/kinematics-correspondences, most clearly afford matching.

I think this difference in agreement between groups of examples, and the way it may reflect sound features affording movement more or less unambiguously, are worthwhile pursuing in follow-up studies. I would propose that future experiments should be designed so that a variety of examples may be examined within each category of sound features. I also believe that it would be advisable to try out other modes of movement registration, since some of the features that change in a sound (e.g. dynamics) are not readily captured in the restricted format of this study. One option would be to use the *Polhemus system*[76] for a three-dimensional registration of body movements. Using this mode of registration would provide more vivid movement material, but the challenges regarding the analytical approach would most likely be greater.

The issue related to the participants' level of training was addressed in the discussion of interrater agreement. Although not clearly demonstrated in this study it is reasonable to believe that the level of training does affect the proportions of observed agreement/disagreement. In the present material it seems that the sound-tracers as well as the raters are too equal in terms of training. So, in a future study I would propose the participants differ more distinctly in terms of musical training.

Finally, the tentative conclusions I have proposed are based on my own interpretations of the drawings and how they are rated. I asked the raters after each session about their experience of the task and the individual examples, but this was carried out unsystematically. Certainly, it would be interesting to follow up the ratings with a more extensive qualitative interviewing of some of the participants.

[76] See www.polhemus.com

Chapter 7. Analysis of correspondences in music and video-recorded dance movements

7.1. Introduction

The purpose of this chapter is to work out a procedure for the analysis of correspondences in music and movement. This will include a discussion of descriptive terms and tools that may be used for this purpose. The procedures have been worked out to be applied to the analysis of music-movement correspondences as they appear in video recordings of people who were asked to move freely to short excerpts of music. These videos were collected at two separate sessions:

- At the first session we used five different musical excerpts and people who had different levels of expertise participated
- At the second session, which I shall refer to as the follow-up session, we invited two trained dancers and chose musical material that was more focussed in terms of certain features (e.g. loudness, the density of onsets, timbre, articulation)

In the next chapter (8) I will concentrate the study on videos made during the initial session, i.e. based on two musical excerpts and their corresponding gestural variants. The musical components were captured from a CD-recording of György Ligeti's *Ten pieces for Wind Quintet* from the movements *Lento* and *Prestissimo leggiero e virtuoso*. Hereafter, the musical excerpts will be referred to as the *Lento* and the *Prestissimo* excerpt. In the qualitative analyses of these recordings I shall concentrate on the recordings of two female dancers that performed three movement variants to each musical excerpt, i.e. six different gestural interpretations of the *Lento* and six different interpretations of the *Prestissimo*. The movement variants will be referred to as the *Lento* and *Prestissimo* variants[77].

The present chapter and the next are thus a continuation of the investigations carried out in chapter 6, but the empirical material differs fundamentally from the sound-tracings:

- The musical component is of a longish time span (15-20 seconds) and is also more complex in terms of the number of co-evolving/interacting features.
- The gestural component is also richer since full body-movement will be analysed.
- The material is characterised by the way the visual and the auditory streams occur as simultaneously co-evolving events and not as separate events as in the sound-tracing study. This implies that a study of correspondences also has to consider any interaction effects as a result of multisensory integration (see chapter 3).

The related notions of *shapes* and *shaping* will be elaborated on in more depth. In the analysis of the sound-tracings in the previous chapter, a drawing on the two-

[77] The variants are available as video-files on the CD, see audio-visual examples *avex1 – avex12* in the folder /audio-visual examples chapter 7+8/. The two dancers who appear on the videos are Åshild Ravndal Salthe and Mona Næss - many thanks to both of them for their contribution.

dimensional tablet was understood as a geometrical representation assumed to reflect how the sound-tracer heard changes in the sound, e.g. changes related to pitch, dynamics or timbre. In other words, shape is connected to the way the perceiver 'picks up on' the way features change over time. Shape will be used so that it can refer both to a geometrical representation (or interpretation) of perceived changes (e.g. a drawing such as in the sound-tracings) and changes in features in the sound. Understood such, the drawings in table 17 reflect how the sound-tracers explored shapes afforded by the sound by means of drawing movements.

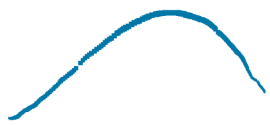
	Example of drawn shape that alludes to the pitch (kinematical) shape of the sound
	Example of drawn shape that alludes to the dynamical shape of the sound
	Example of drawn shape (manner of drawing from even to rougher) that alludes to the timbral shape (timbral changes) of the sound

Table 17. The drawing movements in the sound-tracings are understood as reflecting perceived change in the sounds.

The videoed dance movements will be understood similarly as gestural explorations of evolutions in musical sound. The perceptual process of exploring sound in terms of shapes has been referred to as shaping (see chapter 5). Applying the Gibsonian concept of affordance, I would understand shaping both as something the phenomenon (sound) offers the perceiver, as well as something the perceiver does to the phenomenon (sound). On the one hand we may understand shaping as extracting shapes, but at the same time, within the same perceptual process, as imposing shapes. The perceptual exploration of sound occurs in the nexus of extracting and imposing shapes.

Changes in activation (introduced in chapter 4) represent one such shape that I pre-suppose to be of particular relevance to the analyses of music and body-movement relations. The two examples from Ligeti's *Wind Quintet* differ in terms of overall activation contour. The excerpt from Ligeti's *Lento* is characterised by a change from a low to a slightly higher level of activation, whereas the excerpt from his *Prestissimo* changes from a high level to a lower level of activation. We may imagine the changes as schematically visualised by two different shapes:

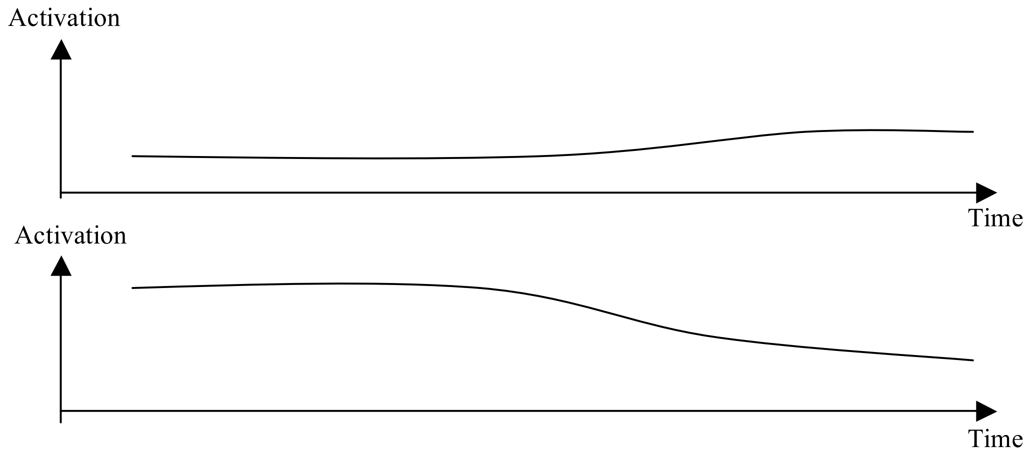


Figure 57. The curves suggest differences in activation contours between the *Lento* and the *Prestissimo*.

The question arises whether the simultaneous gestural flux matches these contours in terms of activation and how the degree of match influences perceived correspondence. Activation is understood as a global emergent quality, and in the present chapter I shall work out two sets of activation features based on the assumption that individual features in music and movement respectively contribute to the overall activation contour. For example, we may assume that the density of onsets from a low to a higher frequency of onsets as in the *Lento* is one of the features that influences the overall contour. Accordingly, changes in *density of onsets* may be understood as a sub-shape.

Similarly, shapes may be connected to features of movement that are approached by Laban Movement Analysis, for example a musical or gestural process may be characterised as a change from 'bound' to 'free', or from 'pressing' to 'gliding'.

Closely connected to the way sensory streams are perceived on the basis of shapes, correspondences will also be discussed with respect to *patterning*, and the way events in some cases are characterised by a structure consisting of *prefix – goal-point – suffix*. In music this has something to do with the way a musical chunk (e.g. a melodic phrase) is temporally centred around a goal-point. The question is how correspondence is influenced by the way the phrase structures of music and movements resemble each other. In this regard, the analysis will take a special interest in the emergence of *synch-points* (moments of simultaneous accentuation in the music and movement, see chapter 5) since such audio-visual peaks are considered to play a prominent role in convergence.

In sum, the following interwoven aspects will be addressed in the analysis:

- The general activation level and overall activation contour
- Descriptions of more subtle movement nuances, in the analytical procedure denoted dynamics (of movement)
- Kinematics
- Chunking and phrasing
- Moments of accentuations in the music and movement, as well as moments of converging accentuation, i.e. synch-points

This list sums up the basic features that I assume contribute to the perception of correspondence. Together, they provide a framework within which the empirical material will be described and discussed. Furthermore, they should be understood in connection to the overall assumption that music and movement correspondences are characterised by being both flexible and non-arbitrary. For example, convergence on the basis of similar activation contours may be viewed as a non-arbitrary aspect of correspondence; but again, the listed general features should not be understood as fixed prerequisites for correspondence to occur. This means that the listed features will be used as a way of observing and describing more in detail how correspondence may be viewed as *flexible*, yet also *non-arbitrary*.

The audio-visual material addressed in this chapter is meant to provide a basis for describing details of music and movement correspondences and to exemplify the fundamental assumption about flexibility/non-arbitrariness. It should be noted that since the observational study did not cover a wide range of musical styles, and did not involve a large number of participants, the basis for drawing firm and general conclusions is limited. However, I believe that the study offers the potential to describe tendencies as well as a basis for discussing and elaborating on theoretical considerations. Methodologically, the analysis is best characterised as an explorative study-approach that attempts to address the material from many different angles and perspectives.

In the first part of the chapter I shall explain the procedure for the collection of empirical material. Issues concerning the features contributing to correspondence will be discussed in detail, as well as how they are implemented in the analytical approach. This includes description of software that will be used for qualitative and computer-based analysis.

7.2. Collection of material: videos of free dance-movements

In an observational study we videoed participants who were asked to make movements that they felt matched musical excerpts. This was carried out in two main variants: air-instrument playing and free dance-movements. Additionally, we tried out different ways of registering and reducing the movement data. At one session participants were asked to move only one arm and they were equipped with a glove with sensors. In another session of recordings of full body-movement, the participants were equipped with coloured gloves to enable movement-tracking. Here I shall concentrate on the recordings of full body-movement in the free dance-movements task.

7.2.1. Musical material and participants

For this part of the study a sound-file with five different musical excerpts typically lasting between 15-20 seconds was prepared. The first three were taken from György Ligeti's *Ten Pieces for Wind Quintet* [78]. The first excerpt was the opening of the third movement of the work, the *Lento* (19 seconds); the second was the opening of the fourth movement, *Prestissimo leggiero e virtuoso* (15 seconds); and the third was the eighth movement, *Allegro con delicatezza*. The fourth example was rendered from a performance of Vivaldi's *The Four Seasons* [79], and the final segment was an excerpt from an electronic composition composed by one of the members of the research group, Alexander Refsum Jensenius[80]. The sound-file was prepared so that each example was played three times in succession with a short break between of two seconds. The sound-file lasted approximately five minutes.

The musical examples were chosen to offer a range of musical features such as changes in dynamics, articulation, tempo, timbre/texture, as well as pitch features (e.g. pitch range). Furthermore, they were characterised by the following:

- Non-periodic (except for the Vivaldi example).
- Not readily associated with a specific dance-style or groove, e.g. a waltz in ¾ metre.
- Differences in levels of activation.
- The excerpts were cut and edited so that they could be heard as a relatively closed musical unit/phrase consisting of one or two chunks

The choices were made so that the task could be relatively open, i.e. to avoid the participants being 'forced' into a periodic rhythmic pattern or a metrical beat-pattern. Music from the Western art music tradition primarily based on timbral/textural and dynamic features was favoured, since our main interest was to observe how the participators would respond to dynamical/kinematical features, and the way they would explore or interprete the continuous stream of musical sound in terms of chunking and phrasing.

7.2.2. Procedure and instructions

The recordings were made against the background of a blue screen and floor[81]. Two high-quality DV-cameras were used, one positioned at the front and one at the side, both cameras at a distance of four metres from the dancer. In addition, firewire web cameras placed in the same positions were used for real-time video analysis[82]. A marker was placed on the floor to indicate the position of the participants.

---

[78] Ligeti, G. (1998). Ten Pieces for Wind Quintet. On *György Ligeti Edition Chamber music* [CD]: Sony Music Entertainment Inc. Performed by London Winds.
[79] From *Winter* movement: *Allegro Non Molto* from Vivaldi's Concerto in F Minor, Op. 8/4 RV 297- Il Giardino Armonico. 1994. "Vivaldi: The Four Seasons." [CD] Teldec.
[80] From *Improvisation* by Alexander Refsum Jensenius (piano and live electronics), Marie Fonneløp (voice), Arve Voldsund (live electronics)
[81] The recordings were made at the *Intermedia* video-studio at the University of Oslo
[82] Musical Gestures Toolbox (MGT), see www.musicalgestures.uio.no (software)

The participants were recorded one-by-one without any other of the participants watching. They were informed about the study, i.e. the general aim of the *Musical Gestures Project*, as well as that the intention of the observational study was to observe and analyse the relations between music and body movements. They also completed a form about their own musical experience (years of formal music education, years of training on a musical instrument, and years of experience with movement-related activities, such as dance/ballet, aerobics and sports).

After this brief introduction, they were told that they were going to listen to a series of musical excerpts: five different examples repeated three times each with a short break between each excerpt. They were then asked to make movements that they felt matched the music. We further explained the task by asking them to move to the music as though they were visualising or sketching the music in the air. It was emphasised that there were no right or wrong responses and that we were interested in their personal and immediate interpretation of the music. We also encouraged them to try out different movement patterns when the examples were repeated, if they "felt like it". They were asked to remain close to the marker on the floor so that the movements were mainly performed by the upper part of the body.

The movements were improvised on the spot; they had not been allowed to listen to the musical examples before the video recordings were made. There were two reasons for recording three variants of each example:

- First, in this initial phase of our work, we wanted to observer differences between the fully improvised version and the succeeding "learnt" variants. However, this aspect will not be in focus here.
- Second, we assumed that a number of variants would enable the differences to be analysed systematically. This will be discussed as a methodological issue later in this chapter.

7.2.3. Participants

A total of six persons were included in the part of the study that used the musical material described above. Three were trained dancers[83], and the others reported that they had had no specific training, either in music or in movement-related activities. The three dancers and one of the novices were video-recorded wearing coloured gloves to enable movement tracking, and they were asked to move the upper part of the body, as explained above. The two remaining novices performed the task by moving only one arm equipped with a sensor-glove. Our first impression was that all the participants, regardless of their prior training, clearly demonstrated sensitivity to music and movement relations. Of course, the dancers were much more comfortable in a situation like this as movement to sound is often their everyday and professional context. Not unexpectedly, they demonstrated a richer and much more nuanced and precise repertoire of movement. However, it was also interesting to observe how the novices responded to variations in tempo and articulation in the music, although their movement was coarser than that of the trained dancers.

[83] Students at the National Academy for Ballet

For illustration, table 18 compares a Quantity of Motion (QoM) analysis of one of the novices and one of the dancers in the selected *Lento* and *Prestissimo*. The QoM-analysis is based on the change over time in pixels in a video image. This means that when there is a high level of movement in the video-recorded body, this leads to peaks in the QoM-curve (vertical axis), and when the movement is fast and abrupt this will be shown on the curve as sharp peaks. The QoM-analysis is here used as a coarse visualisation/representation of changes in activity, i.e. an analysis that indicates the activation profile of a process of movement (see further explanation of QoM-analysis in the next section).

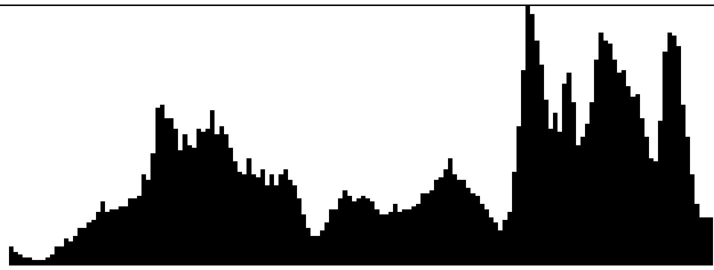
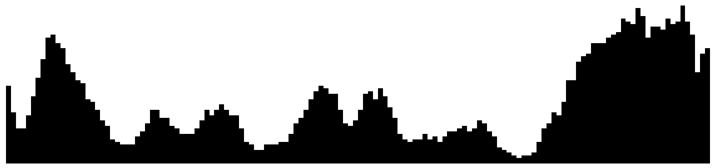
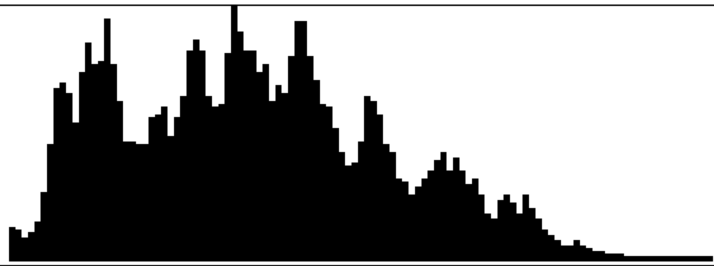
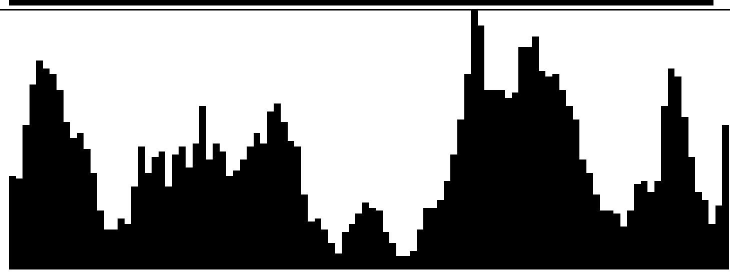
Lento	Expert	
	Novice	
Prestissimo	Expert	
	Novice	

Table 18. QoM-analysis of expert vs novice variants.

The analysis of the expert variants suggests the following:

- *Lento*: a change from low to a higher level of activity, a rounded curve suggests smoothly articulated movements
- *Prestissimo*: a change from high to a lower level of activity; more abrupt curves suggest quick movements corresponding to the music

The QoM-analyses of the novices' interpretations are not as clearly articulated with respect to the contour compared to the experts' curves. But if we look more closely at the curves we see some of the same tendencies: the rounded curves in the *Lento* suggest slower and more smoothly articulated movements compared to the sharp peaks that characterise the *Prestissimo*.

The reason for including participants who had different levels of training was that we wanted to observe whether experience affected the way participants responded, and thus to form a basis for evaluating theoretical assumptions about the role of motor imagery in music perception. This is a theme that I shall not pursue in this analysis. I shall use the material from a solely qualitative perspective, i.e. to describe variations in music- movement correspondences. For this reason, I will concentrate on the variants performed by the trained dancers since they provide more vivid material. However, in any future analysis it would be interesting to see if there were systematic differences in the way novices and experts approached a task like this, for example with the hypothesis that novices might tend to respond to apparent correspondences such as one between a rising pitch contour and an upward movement (kinematical shapes), rather than responding with more subtle nuances in movement character.

7.2.4. Follow-up session

The observational study described above was carried out at an early stage of our work. When we made the videos, analytical concepts and tools had not yet been fully worked out. Later in the process we saw the need to work with less complex musical material to facilitate systematic analysis. As the notion of activation as a theoretical and analytical concept was elaborated on, a set of *activation features*, (see later in this chapter) was used as a basis for selecting the musical material for a follow-up session of video recordings. The idea of this follow-up recording session was to use the same procedures and instructions as in the first session to make videos. The musical excerpts were selected from CD-recordings with the same general requirements (non-periodic, texture/timbre-based rather than traditional melody/harmony-based etc.). Attempts were made to find examples that were relatively low on dimensionality, i.e. so that the musical flux was mainly characterised by changes in a limited number of features (e.g. density of onsets or articulation). We wanted to observe in a more controlled manner whether changes in one musical feature affected changes in specific gestural features, e.g. whether the density of onsets in music consistently affected the density of onsets in movements.

Two sound-series with ten musical excerpts in each series, and in which each excerpt was repeated twice, were prepared. The material collected in this session will not be commented on in detail, but a few selected examples will be used to illustrate the idea of *activation features*, to be explained later in this chapter[84].

[84] These examples are available as mp4-files on CD-rom, see audio-visual examples *avex13 – avex19*

7.3. Analytical tools

To aid analysis I have utilised tools that in different ways visualise aspects that characterise the co-evolving musical and gestural processes. These are briefly listed and described in the following:

7.3.1. Anvil

Anvil is a tool designed for the manual/qualitative annotation of gestural processes. Michael Kipp, a researcher within the field of non-verbal communication, developed the software to facilitate the detailed analysis of co-verbal gestures (Kipp, 2004)[85]. The tool facilitates a precise description of changes along a timeline and with respect to predefined observational categories. For my purposes I have worked out a set of observational categories which has been written into a specification file (in xml-format). Each observational category is recognised as a track in the annotation and arranged within groups of tracks. The annotation board includes four groups of observational categories: general, activation, effort and correspondence:

- General
    - Chunks
    - Kinematics
    - Dynamics
- Activation
    - Density
    - Extension
    - Involvement
    - Speed
    - Force
    - Articulation
    - Emphasis
- Effort
    - Time
    - Weight
    - Space
    - Flow
- Correspondence
    - Degree
    - Synch-points
    - See-hear
    - Not see-hear

Table 19. The Anvil annotation board enables manual annotation according to four main groups of observational categories: general, activation, effort and correspondence.

Observing each movement by concentrating on and describing one category at a time has been very useful in the working process as the annotation provides a descriptive overview of each gestural variant. By default the annotation board also

[85] See also: http://www.dfki.de/~kipp/anvil/

includes a waveform of the sound component. Figure 58 shows an example from the annotation board of one of the *Lento* variants performed by dancer B:

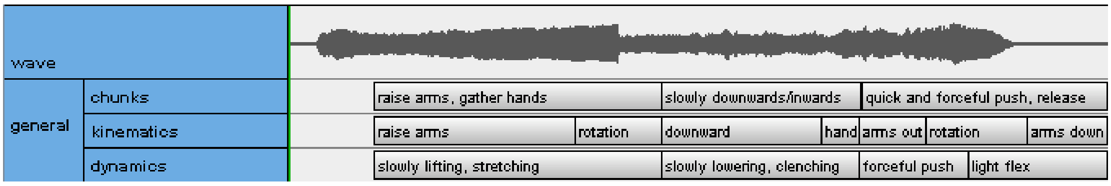


Figure 58. Example of annotation board including waveform of *Lento*excerpt and annotation of chunks, as well as kinematics and dynamics of one of the movement variants.

Some of the categories have already been discussed theoretically in previous chapters. The categories and the use of Anvil will be further explained later in this chapter with respect to the way they are applied in analysis.

7.3.2. Musical Gesture Toolbox

One of the participants in the Musical Gestures project, Alexander Refsum Jensenius, has developed a set of analytical tools based on the Max/MSP/Jitter environment[86]. The Musical Gesture Toolbox includes modules for the analysis of both movement and sound (Jensenius, 2006; Jensenius et al., 2005)[87]. To aid the analysis of movement variants, I shall use the following tools:

Quantity of Motion is an analysis of a video stream that is based on the amount of changes in all pixels of a video image from frame to frame. The calculation is made by adding up the active pixels in a movement image and plotting the value over time (Jensenius, 2006). Thus, the analysis provides a coarse indication of the quantity of activity and how this activity changes over time. However, it does not say anything about which body parts are active. In my analysis I shall regard Quantity of Motion (QoM) as a tentative indication of changes in activation, and it will be used as such in combination with the manual annotation of activation features in Anvil.

The Max-patch I have used to create the QoM-analyses uses an auto-scaling function[88]. This means that during video-playback, the patch automatically scales the values on the Y-axis (showing the QoM level) so that the curve is vertically stretched out. For each video the patch ‘learns” the amount of scaling required to fill out the Y-axis. This means that the QoM-curves show the relative changes in quantity of motion over time (X-axis) within one movement variant. The QoM-curves used to analyse the movement variants have been obtained by running one movement variant at a time through the patch so that each QoM-analysis used a different level of scaling for each variant. This means that it is not possible to compare absolute levels of QoM between

---

[86] Cycling ’74. Max/MSP 4.6 Jitter 1.6. Graphical audio and video environment (computer program). http://www.cycling74.com

[87] For further information see: http://musicalgestures.uio.no

[88] See Jensenius, 2008; pp 148-149 for a description of the auto-scaling function.

the variants but the analyses are useful to show differences in overall contours between the variants.

Motion history images. To visualise in one single image the trajectory of a movement, the patch called motion-history-images1.mxt has been developed (Jensenius, 2008). It has been developed on the basis of an 'open-shutter' technique and uses different types of video-feedback algorithms. The image in figure 59 shows the trajectory of a slow upwards movement performed by one of the dancers. This kind of representation is particularly useful to visualise the kinematical aspects of a movement.

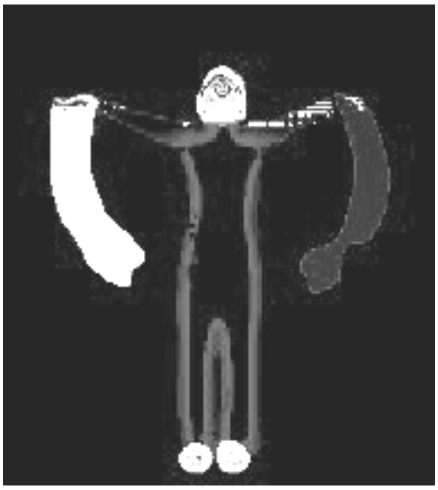


Figure 59. Motion history images show the trajectory of a movement.

7.3.3. Signal-based musical sound analysis

For analysis of the musical component I shall refer to signal-based analyses such as *waveform representation* and *spectrogram*. A waveform representation shows variations in amplitude as a function of time and provides a rough image of the way the dynamical level evolves within a time window (McAdams, Stephen, Depalle, & Clarke, 2004). For my purposes, a graphical representation of this kind is useful to illustrate shaping characteristics and to refer to beginnings/endings of chunks (see *Varese* example in chapter 5, figures 20 and 21). Similarly, a spectrogram, i.e. a representation of how frequency components are present in the sound, their strength, and how they evolve over time, may visualise structural features as well as the distribution of tone/chord onsets. Additionally, the spectrogram provides a basis for describing timbral characteristics, e.g. in the way the spectrum indicates formant structures, spectral characteristics related to perceived loudness, as well as the time-varying spectral envelope (Roads, 1996; p 544). In my discussion, with a focus on changes in activation, the spectrogram will be used to visualise tendencies related to brightness/loudness, i.e. that strong frequency components in a certain high register may in many cases be perceived as having a higher degree of brightness which in some cases may be associated with a higher level of activation.

The signal-based tools for analysis will primarily be used to visualise the way features of music and movement change over time. They are meant to be tentative and relative indications of changes within a time window and will be used in combination with metaphorical/language-based descriptions. They are not understood as absolute measurements of musical and gestural signals; e.g. as explained above, a QoM-analysis of one gestural variant is not directly comparable with the same analysis of another variant. However, they are useful in the way they visualise tendencies within each single variant (with respect to the overall contour) as well as in the way they visualise similarities/differences in overall contour between the variants.

7.4. The free dance-movements material compared to the sound-tracings and air-piano collections

Clearly, the analysis of this material is challenging in another manner compared to the sound-tracings that were discussed in chapter 6. First, the sonic excerpts were simple and short in duration, ranging from single sounds to single phrases, and they never exceeded the limits of one sonic Gestalt or chunk. The excerpts from the *Lento* and *Prestissimo* movements are longer in duration, 20 and 15 seconds respectively. Research into memory points in the direction that we have but a limited capacity with regard to how long in duration an event can be and still be perceived as one single chunk (Snyder, 2000), this time window being referred to as our *perceptual present* (Stern, 2004). Based on a review of several investigations of this time-span, Clarke suggests that in music perception the duration of this time window seems to be somewhere between 2 and 8 seconds (Clarke, 1999). If the event exceeds the maximum duration we tend to break it up into more than one chunk. This implies that the *Lento* and *Prestissimo* excerpts most likely are perceived as consisting of more than one chunk. Whereas the sonic segments in the sound-tracing study were very clearly defined, the phrase structure, for example in the Lento, is comparatively vaguely suggested. Additionally, the excerpts, with for example simultaneous changes in timbre, dynamics, articulation and pitch, all demonstrate the complexity and multidimensionality that characterise real, composed music. The longer time-span and the higher degree of dimensionality mean that the analysis has to deal with the problem that it may be difficult to identify exactly the features in the music that the dancer responds to within a given temporal window. This is much clearer in the sound-tracing study.

Second, in the sound-tracing study the visual movement is presented as a frozen image; moreover, it is reduced to a two-dimensional representation within a very limited space. Obviously, this reduction facilitates comparison. In the dance-study we have two simultaneously evolving streams. The problem of describing musical sound with respect to temporal characteristics is well known in music analysis. In the process of working out analytical procedures it has emerged that we face a similar challenge when attempting to describe the changes in features in continuous streams of movement as well. This means that in analysing correspondences we have to deal with two temporal streams that because of their transitory features are not easily represented in a mode that enables comparison. In addition to this problem of description and representation, we may assume that interaction effects occur when the two streams are combined, i.e. if a convincing method for the representation of the musical and the gestural signal were available, these representations might have been to some extent insufficient due to interaction effects.

Our observational study included video-recordings of different kinds of air-playing, such as air-piano, air-guitar, and air-drums, in addition to the free gesturing to music. Typically, the participants were asked to listen to audio files with pre-recorded musical excerpts played on the piano, guitar and drums respectively, and were supposed on the spot to play in the air as though they were playing the instrument.

The recordings of air-piano playing were further analysed with respect to music-movement correspondences (for details on musical material, participants and qualitative judgement, see Godøy et al., 2006b). The purpose was to observe the relationship between air-piano movements and the corresponding piano-music excerpts. Five subjects were chosen according to their different levels of musical training, ranging from one novice to three intermediate players and lastly one professional pianist, the expert. It was assumed that the expert would perform better, i.e. the degree of correspondence would be higher compared to the novice. We also predicted that the novice and intermediate players would perform relatively better on global levels of correspondences and less well on details. These levels of correspondences, starting with overall activity correspondence, followed by pitch, onset, dynamics and articulation correspondences were explicitly defined and used as a basis for the observation and evaluation of each performance. The seven different levels, or observational categories, were defined as follows:

1. Overall activity correspondence, i.e. density of movements in relation to density of onsets in the music, but regardless of pitch and onset precision
2. Coarse pitch-space/keyboard-space correspondence, i.e. relative locations of hands left-to-right on an imagined keyboard at phrase/section-level
3. Detail pitch-space/keyboard-space correspondence, i.e. relative locations of fingers on an imagined keyboard at note-by-note level
4. Coarse onset correspondence, i.e. synchrony at down-beat or event level (event instead of downbeat in cases of less or non-periodic music)
5. Detailed onset correspondence, i.e. synchrony of finger- and/or hand-movements at note-to-note level
6. Dynamics correspondence, i.e. size and speed of hands/arms/body movements in relation to loudness
7. Articulation correspondence, i.e. movements for accents, *staccato*, *legato*, etc.

Correspondence using five musical examples was evaluated for each participant. Three members of the research group[89] conducted the qualitative judgement, i.e. the raters studied the audio-visual material and agreed on a score for each variant according to an ordered categorical scale:

- No correspondence, score = 0
- Poor correspondence, score = 1

[89] Rolf Inge Godøy, Alexander Refsum Jensenius, Egil Haga

- Approximate correspondence, score = 2
- Good correspondence, score = 3

For illustration we estimated mean values for each participant and this pointed in the direction that the experienced musicians performed systematically with a higher degree of correspondence. Considering the low sample size with respect to the number of participants, no firm conclusions were drawn regarding the role of training.

The point to be made here concerns the basis on which correspondence was judged. As raters we based our judgements on our personal (ecological) knowledge of piano-playing, i.e. asking the following question for each variant and addressing each observational category from overall (e.g. overall activity) to detailed (e.g. articulation) correspondence: "Do the movements of the air player correspond to the way I imagine the piano sound has been produced?"

This means that judging correspondence in the air-playing material is a comparatively well-defined task. The keyboard and imagery of the required sound-producing actions provide a point of reference on which judging correspondence is based which also makes analysis more straightforward. However, in comparison the openness that characterises the material from the free dance-movements implies that we do not have this solid point of reference. This makes the analysis of correspondences in this material more open-ended compared to the analysis that was carried out in the air-piano study.

7.5. Elaboration of activation concept

I would propose that the observational categories worked out for the air-piano study may serve as a model for a similar approach to the analysis of music and movement correspondences in the task involving free dance-movements. The activation contour of music and movement respectively may serve as one overall point of reference in judging correspondence. In the next sections I shall work out two sets of observational categories, one for music and one for movement. These categories will be referred to as *activation features*. They are understood as separate features which together contribute to the overall activation contour of music and movement. With respect to the separate activation features, music and movement may correspond in two ways:

- Indirectly, in the way the separate activation features contribute to the emergence of a global activation contour
- Directly, in the way a specific activation feature in music converges specifically with an activation feature in movement; for example, we may assume that the density of onsets in the music corresponds to the density of body movements.

7.5.1. Activation features in music

I understand the activation contour of a musical excerpt to be related to the way the music affects the body. While listening to music it seems that music to varying degrees evokes different levels of energy, either in one's imagination, or, if allowed to move, the body actual moves with different levels of activity or activation.

Occasionally, the imagined or actual movements are forceful, extended in space and spread fully throughout the body. In other cases the movements are more restricted in space and are more gently and smoothly performed. The two musical excerpts that are the focus of analysis in this chapter demonstrate different levels of activation, i.e. the *Lento* has a low level of activation and the *Prestissimo* has a comparatively high level of activation. Moreover, there are fluctuations within each excerpt. These temporal variations in activation evoked by the music, i.e. the imagery of motor activity, make up the activation contour perceived.

But how can activation in music be described and analysed with reference to musical characteristics? In other words, which musical features contribute to the way motor imagery is evoked, thus resulting in an overall activation contour?

As discussed in chapter 4, music theory acknowledges that perceived *intensity* or *energy* in music is affected by a number of co-occurring and co-evolving musical features, such as dynamics (loudness), timbre, articulation and density. Similarly, I would understand activation as a global, emergent quality resulting from many contributing features. As also explained in chapter 4, the reason for choosing the term activation rather than intensity or energy is that it more readily alludes to the motor imagery that is evoked while listening to music. I would suggest that the following set of musical activation features may contribute to overall activation:

- *Density of events*, i.e. a higher density (more events in time) leads to a higher level of activation. Events is used here to mean the onsets of chords and tones.
- *Pitch variations*; i.e. relatively large variations in pitch result in a higher level of activation
- *Loudness*[90], i.e. loud music (e.g. *fortissimo*) results in a higher level of activation.
- *Articulation*, i.e. accentuated/distinct articulation is thought to result in a higher level of activation than music that is played more *legato*.
- *Timbre*, i.e. a bright timbre may be perceived with a higher level of activation than a dull timbre.
- *Texture*; i.e. a horizontally spread and thick texture may result in a higher level of activation than a more concentrated, thin texture.

In the following I shall discuss each of these individual features in connection with motor theory. For illustration I shall refer to examples from the follow-up free dance-movement study where the musical excerpts were selected to enable the observation of how specific features, e.g. the density of onsets, might afford specific features of body movements, i.e. whether an increased density of onsets would lead to a general increased level of activation and whether this would be specifically articulated as an increased density in body movements.

I regard the listed activation features and the way each of them is assumed to affect the overall activation contour as tentative hypotheses about music-movement relations. For example, an increased density of onsets may lead to a sense of increased activation in music, thus affording an increased level of activation in the movement response; or, a musical passage that is distinctly played may lead to a sense of

[90] I prefer the term *loudness* rather than *dynamics* since the latter will be used as descriptive term in movement analysis later in the chapter

increased activation in music, thus affording an increased level of activation in the movement response. Examples from the follow-up study have been selected to illustrate the proposed link between musical activation features and movement activation features. I have chosen the variants which I think best exemplify this link.

7.5.2. Density of onsets

We might think of density of onsets in time as a feature that is directly associated with the overall tempo in the music. However, understood within the framework of motor theory, onsets are not weighted similarly in perceptual experience. For example, grace notes performed with a relative high density, but within an overall slow tempo in metre-based music, do not necessarily result in an increased sense of activation as they are perceived as co-articulations, and thus subordinate to a larger and more extended movement unit. In motor theory, e.g. with reference to the action of stretching out the arm to pick up an object, co-articulation refers to the sub-movements in the hand (shaping the hand) and fingers (preparing the grip) that are performed simultaneously with the overall stretching out movement of the arm (Rosenbaum, 1991). Applied to playing musical instruments this implies that grace notes or other kinds of relatively rapid passages, for example in piano-playing, are performed with one overarching arm/hand movement (with a low density in time) co-articulated with many small finger movements (with a high density in time). In light of this I would propose that the density mainly influences the overall activation in time with the overarching arm movement as a point of reference.

For illustration, we may consider three examples. The first two are examples of non-metrical music. In an opening passage of György Ligeti's *String Quartet No. 2* [91] the violins play pizzicato with an increasing onset density which towards the end of the excerpts is heard as a dense texture of light plucking. This is illustrated in the waveform representation and spectrogram in figure 60:

[91] Ligeti, G. (1988). String Quartet No 2 [CD]. Performers: Arditti Quartet. 3rd Movement: *Come un meccanismo di precisione* (0:00 – 0:25 on track 8)

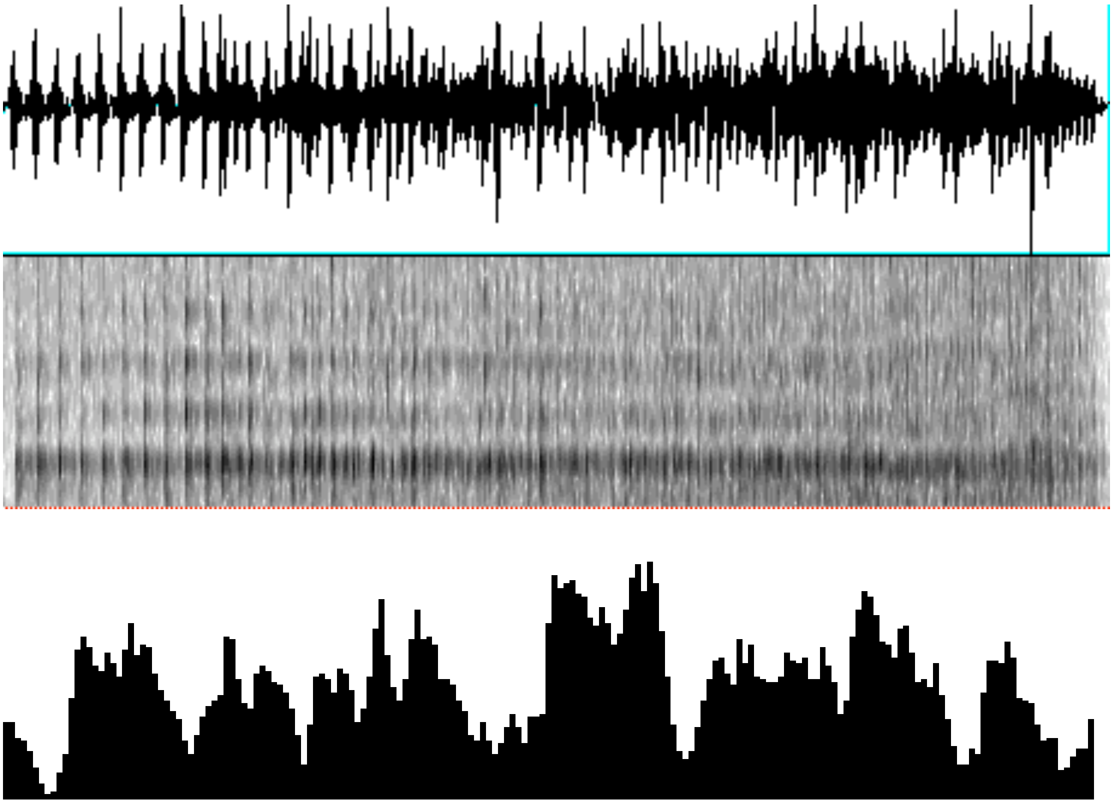


Figure 60. Waveform and spectrogram of a passage of increasing density of violin plucking compared to a QoM-analysis of a corresponding movement performance.

The QoM-analysis suggests that the dancer is responding with slightly increased activation towards the middle of the segment[92]. It should be noted in this regard that although the increased density of onsets might afford an increase in activation from the beginning of the segment to the end, the articulation of the plucking changes from quite distinct at the beginning to a lighter quality so that the texture appears more transparent and shimmering towards the end. Thus, the change in articulation/texture might lead to a sense of decreasing activation which means that changes in density and articulation are working in opposite directions.

In the second example, from Ligeti's *Six Bagatelles for Wind Quintet*[93], the onset density starts at a relatively high level but decreases markedly towards the final tone of the excerpt. The dancer responds clearly to this with a decreasing density in movement onsets which is reflected in a decreasing level of quantity of motion[94]. This correspondence is illustrated in the score[95] and the QoM-analysis (the 'slope' of the QoM-curve corresponds to the figures in the flute and clarinet), see figure 61:

[92] See audio-visual example *avex13* on CD
[93] Ligeti, G. (1998). Six Bagatelles for Wind Quintet [CD]: *György Ligeti Edition Chamber music* Sony Music Entertainment Inc. Performers: London Winds. Movement: *Adagio-Mesto* (1:44 – 2:01 on track 19)
[94] See audio-visual example *avex14* on CD
[95] From Ligeti, G. (1953). Six Bagatelles for Wind Quintet [Studien-Partitur]: Schott.

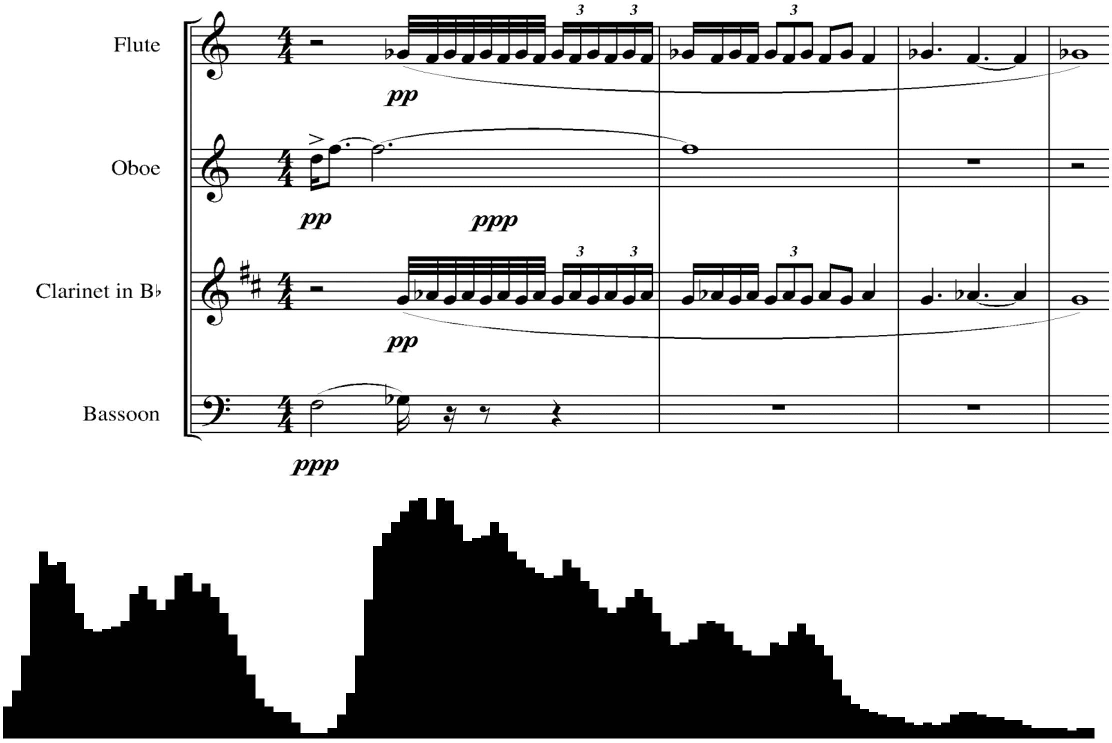


Figure 61. The musical score suggests a decrease in the density of onsets in the clarinet and flute which corresponds to the decrease in activation indicated by the QoM-analysis of the movement performance.

The third example is from the opening of Igor Stravinsky's *Rite of Spring* [96]. Here, it seems that the dancer, at least at the beginning of the phrase, is basing her interpretation on the slow tempo of the vaguely suggested pulse. Towards the end, however, the activity increases as she responds to the increased onset density in the music. Again, the correspondence is illustrated by the score[97] and a gradual increase in QoM:

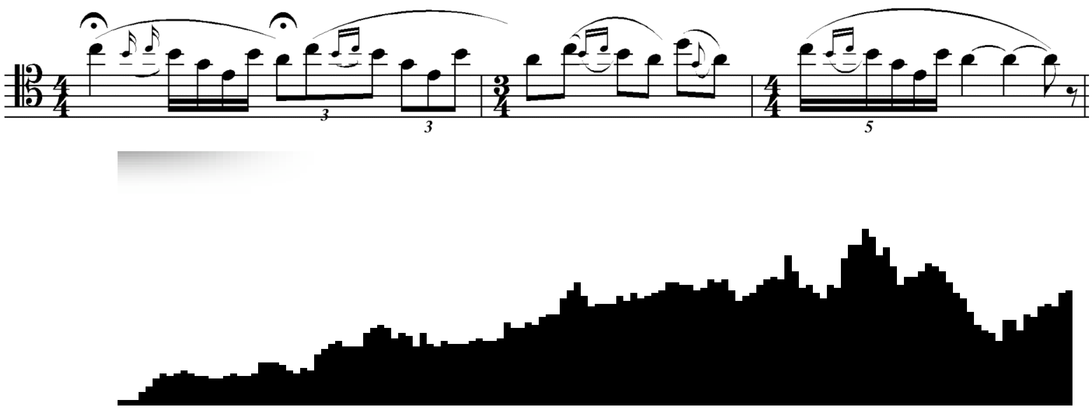


Figure 62. The gradual increase in onset density corresponds to an increase in activation, as suggested by the QoM-analysis.

[96] Stravinsky, I. (1992). The Rite of Spring [CD]. Performers: The Cleveland Orchestra/Pierre Boulez. See audio-visual example *avex15* on CD
[97] From Stravinsky, I. (1967). The Rite of Spring [Pocket score]: Boosey & Hawkes.

7.5.3. Pitch features

As demonstrated in the analysis of sound-tracings in the previous chapter, pitch contours and the trajectory in space appeared to be relatively strong kinds of correspondence. This was also shown in Lipscom and Kim's study of simple sound syntheses and geometric figures, referred to in chapter 3. In chapter 4 I discussed how these observations may be understood from two perspectives:

- as a kinematical aspect, i.e. what moves upwards in physical space tends to be viewed as similar to a sonic pitch contour that ascends in an imagined pitch space.
- as a dynamical aspect, i.e. with reference to motor theory: The tonal displacement between pitch registers requires a motor action. This means that the vertical tonal displacement is associated with the way forces (sound-production) are applied over time to produce the sound change.

One of the examples from the videos clearly illustrates correspondence in terms of the kinematics of music and movement. In this example from Ligeti's *String Quartet No. 1*[98] the musical excerpt was characterised by recurrent glissandi between the low and high registers. This pattern is vaguely suggested in the spectrogram in figure 63:

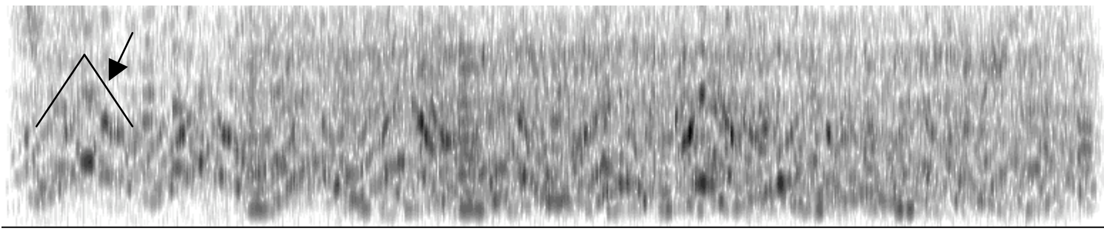


Figure 63. The sawtooth-shape suggests how darker shades in the spectrogram illustrate ascending/descending glissandi in the sound.

The dancer responded to these kinematical shapes with repeated up/down movements (one of these movements is illustrated in the series of stills in figure 64[99]). The up-down contours in the music and the up-down trajectory in the movements suggest a kinematical correspondence.

Figure 64. The dancer responds to the kinematical shape in the music with an up-down trajectory.

[98] Ligeti, G. (1988). String Quartet No 1 [CD]: Mainz: Wergo. Performers: Arditti Quartet. (18:30 – 18:45 on track 1)
[99] See audio-visual example *avex16* on CD

The correspondence may also be interpreted from a dynamical perspective: when foregrounded tonal elements leap between pitch registers and when these leaps are big and relatively abrupt, this may be associated with a higher level of activation. The opening of the *Rite of Spring* exemplifies the relation between pitch features (i.e. pitch range) and the level of activation. The increase in activation observed in the dancer's interpretation seems to be a response to the way the melody makes more extended leaps towards the end of the segment. Thus, there seems to be an interplay between *pitch features* and the *density of onsets*. This is demonstrated even more clearly in the *Prestissimo* from Ligeti's *Ten Pieces for Wind Quintet*, which was introduced at the beginning of this chapter. The representation of variations in pitch range[100] and the QoM-analysis of one of the video-recorded dance-variants (figure 65) exemplify how a decreasing pitch range may result in a sense of a decreasing level of activation.

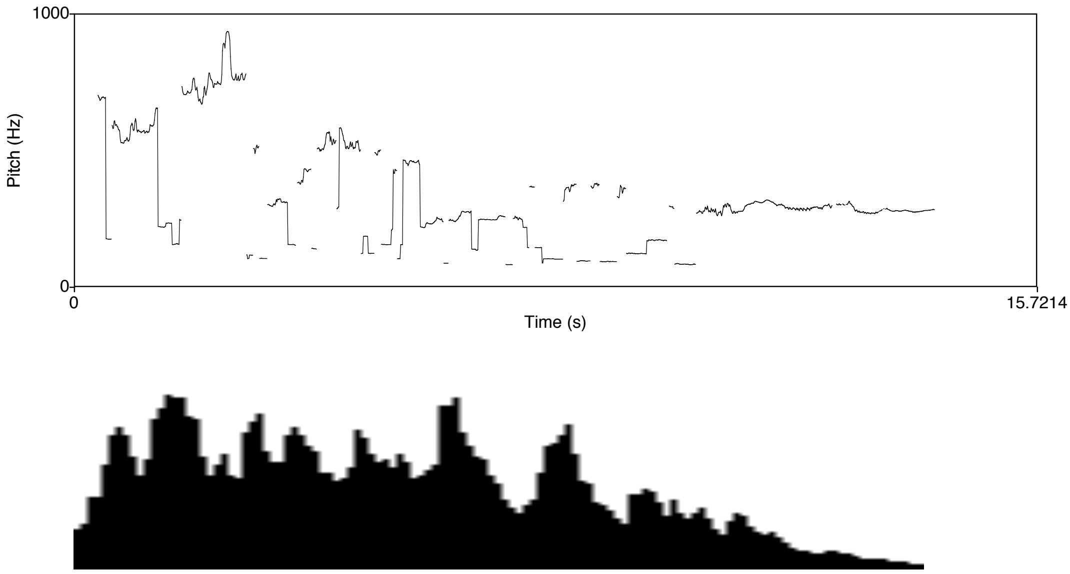


Figure 65. An extract illustrating pitch contour shows how the pitch range between high and lower tones changes from the beginning towards the end. The change in pitch space seems to correspond to a decrease in activation in the movements, as suggested by the QoM-analysis.

7.5.4. Loudness

Initially, dynamics in music have been understood in the usual musical sense, i.e. as changes in loudness (*crescendo* vs *decrescendo; piano* vs *forte*, etc.). In addition, I would suggest that it is useful in this particular context to view dynamics in the light of motor imagery, i.e. that soft music is played with a limited amount of force, whereas *forte* passages are performed with comparatively large and forceful movements. This is in line with the analogous observational category used in the air-piano study, in which correspondence in terms of dynamics was connected to an

[100] In figure 65 the upper display shows a pitch contour extracted by using the *Praat* software, see www.praat.org . The pitch extraction functionality is based on identifying the relative distance between the partials and applying an algorithm to calculate the position of the fundamental, see www.musicology.nl/WM/research/praat_musicologists.html .

imagery of the "size and speed of the arms/hands/body movements in relation to loudness".

The excerpt from Varese's *Amerique*[101] referred to in chapter 5 (see figure 66) illustrates that the recurrent *crescendo - decrescendo* patterns may be experienced on the basis of changes in the way force is applied over time. The example was used in the follow-up session[102], and the QoM-analysis demonstrates how changes in loudness in the music, represented in a waveform display, correspond coarsely to the dance movement's activity peaks, although the movement peaks are not perfectly synchronised with the music peaks in the two first outbursts.

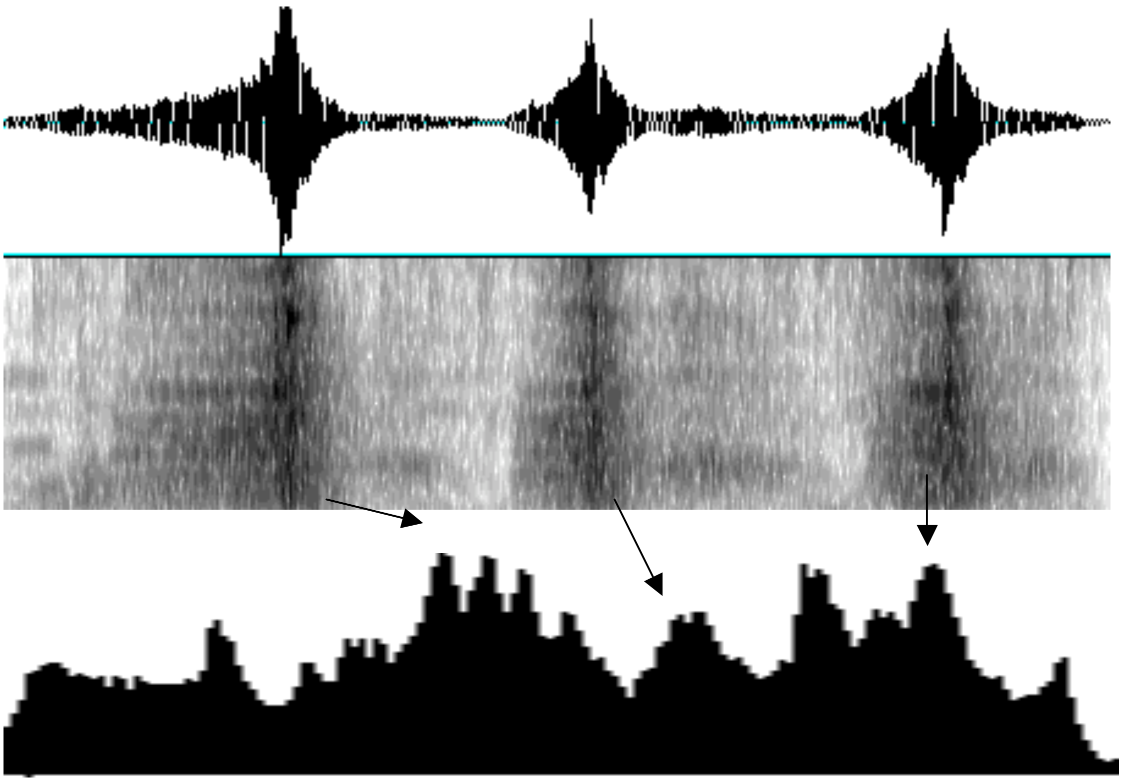


Figure 66. Spectrogram and waveform display from *Amerique*, together with a QoM-analysis of dance movements

In another example, an excerpt from Tchaikovsky'ss *5th Symphony* (the opening of the second movement)[103], we have a similar dynamical shape, but the *crescendo - decrescendo* pattern evolves more gradually. The QoM-analysis suggests that the dancer responds to the more stretched-out dynamical shape: from the beginning of the segment the QoM-curve illustrates that the overall level of activity gradually increases. Additionally, the rounded-off shape of the curve also points in the direction of a relatively low level of activation. Conversely, there is a peak that is more abrupt

[101] Varèse, E. (1994). Amerique [CD]: Decca. Performers: Royal Concertgebouw Orchestra, conducted by Riccardo Chailly (19:00 – 19:20 on CD1 track 2)
[102] See audio-visual example *avex17* on CD
[103] Tchaikovsky, P. (1988). Symphony No. 5 in E Minor, Op 64 [CD]: Essex: Chandos. Performers: Oslo Philharmonic Orchestra, Mariss Jansons. 2nd Movement: *Andante cantabile, con alcuna*, etc. (0:25 – 0:45)

and this moment of increased activation corresponds to and is synchronised with the goal-point in the music, indicated with an arrow in figure 67[104].

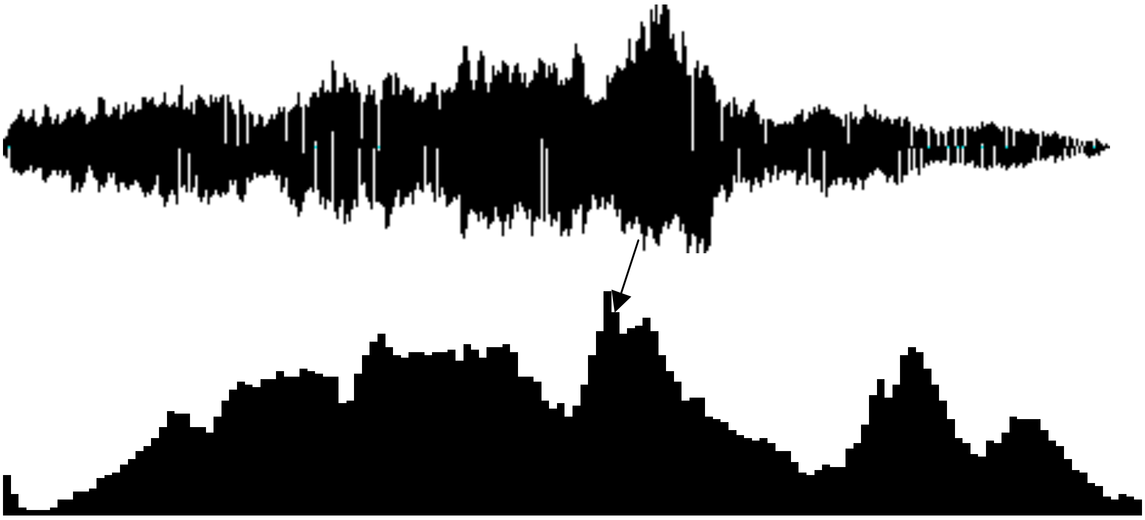


Figure 67. A gradual crescendo in the music corresponds to an increased activation, illustrated by the QoM-analysis. The arrow suggests the simultaneous timing of the goal-point in music and movement.

7.5.5. Articulation

I would suggest that the perception of shadings of articulation is connected to an imagery of the abruptness (*staccato*) or smoothness (*legato*) of the sound-producing action. Articulation is closely related to dynamics in the way the feature has something to do with the size and speed of the imagined movement, but the perception of articulation is centred more on the attack point, or more precisely, on an imagery of the movement that precedes and succeeds the attack. In light of this I would assume that a tone attack that is produced with a relatively forceful and abrupt movement is experienced with a higher level of activation compared to a nuance of articulation brought out by a calmer and more sustained movement.

The opening from *Rite of Spring* exemplifies a legato-style in the musical performance that is interpreted with smooth movements by the dancer. Another example illustrates the opposite kind of articulation. An excerpt from Ligeti's *String Quartet No. 2* [105] is characterised by pizzicato playing which becomes increasingly brutal; it is as though the musicians were pulling the string forcefully and abruptly releasing it, rather than gently plucking the string. The dancer interprets this articulation with jerky, forceful and abrupt movements[106] (see figure 68).

[104] See audio-visual example *avex18* on CD
[105] Ligeti, G. (1968). String Quartet No 2 [CD]. Performers: Arditti Quartet. 3rd Movement. *Come un meccanismo di precisione* (0:25 – 0:45 on track 8)
[106] See audio-visual example *avex19* on CD

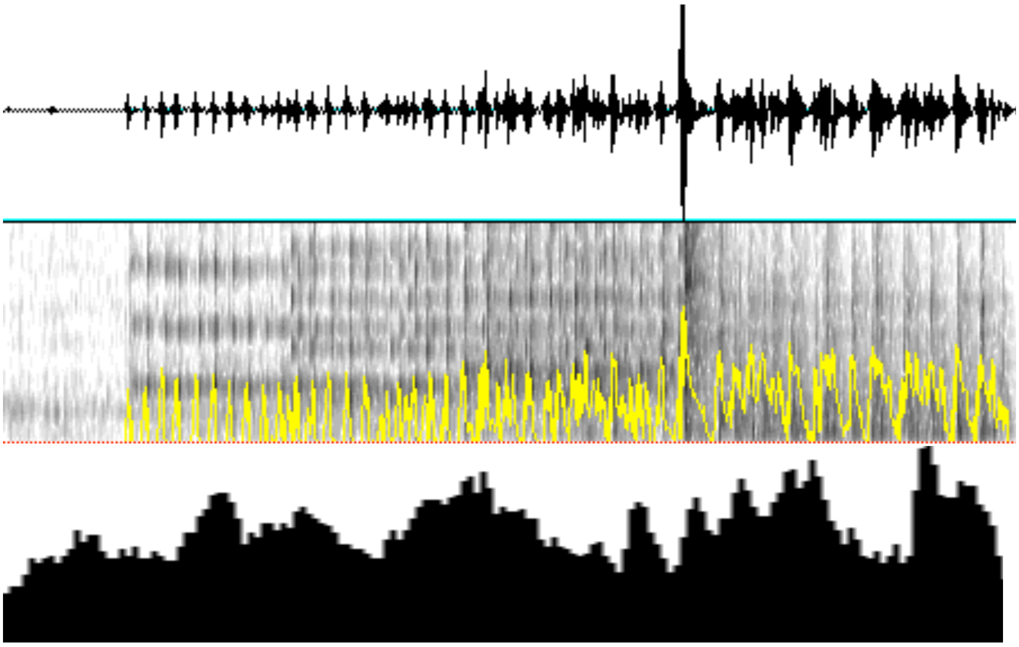


Figure 68. Analyses of excerpt from Ligeti's *String Quartet No, 2*, displayed together with a QoM-analysis of dance movements.

The comparison of the analyses of music (spectrum, intensity and waveform) and movements (QoM) also suggests that there is a simultaneous and corresponding increase in activation, i.e. the level of the quantity of motion gradually increases (see figure 68 above). The peaks of the curve also become more abrupt, which suggest both an increased sense of activation and more distinctly articulated movements.

7.5.6. Timbre

It seems that changes in timbre, for example from dull to bright, would offer increased activation since increased brightness in many cases may be associated with a higher level of intensity. However, timbre appears to be a much less graspable feature within this context, probably because timbre should be understood as a global, emergent feature influenced by those features already discussed such as dynamics, pitch and articulation, and orchestration/instrumentation, harmony/melody and texture. Because of this overarching and highly multidimensional character it might have been fruitful to approach music-movement correspondences with timbre as a point of departure (for a discussion that offers timbre-focussed analyses of the material collected in our first observational study of free dance movements see Casciato, 2005).

The multi-faceted character of timbre is demonstrated by the way the feature may be described in terms of a number of axes or shapes, in addition to the dull/bright axis (for a detailed discussion on timbral shapes, see (Godøy, 1999). For example:

- Open/hollow – narrow/compressed
- Heavy/strong – light/soft
- Dark/light
- Rough/smooth

- Wet/dry
- Warm/cold

The terms are distinctly metaphorical in character since they have been transferred from other domains, such as temperature, lighting, weight, surface features and space, in order to describe timbre.

The shapes also allude to movement to different degrees. For example, we may assume that a heavy timbre is associated with a relatively high level of activation, i.e. the heaviness is connected to an imagery of forceful and weighted movements. I shall not discuss any possible links between timbre and body movement in detail, though I believe that it is relevant to be aware of such connections in music-movement relations. For example, the opening of the *Lento* is characterised by a sustained, narrow timbre; it is as though something were pressing through a narrow passage. It seems that in some of the dance-variants the dancer is capturing this timbral quality with movements that are performed with a controlled, sustained and withheld tension, which may be described by metaphors alluding to movement features such as 'shoving', 'pressing', 'pushing' (to be discussed further in chapter 8).

7.5.7. Texture

Musical texture refers to the relationship between vocal or instrumental parts in a piece of music. For example, voices may be distributed vertically so that they are spread between the high and low pitch register, or the voices may be concentrated within a more limited pitch range. When the music changes from a spread to a more concentrated texture, the music is characterised by this. In this context, texture will be understood in relation to the way textural shapes are used as a prominent musical parameter, e.g. in Ligeti's music (Godøy, 1999). This means that textural changes are viewed as significant in terms of music-movement correspondences when the texture may be described as *spread* vs *concentrated* and/or *dense* vs *thin*. I assume that a spread, dense texture contributes to a higher level of activation than a concentrated, thin texture. Thus, texture is understood in close relation to the pitch features discussed above.

7.5.8. Comments: overlaps and direct/indirect correspondences

The discussion above of density, pitch features, dynamics, articulation, timbre and texture demonstrates that such features overlap. I have commented on the close relationship between dynamics and articulation, density and pitch space, and between pitch space and texture, as well as the way timbre is influenced by a number of interacting features. This means that when using real musical examples that are characterised by co-evolving and interacting features, it is very difficult, if not impossible, to systematically investigate them concentrating on one feature at a time, or observe how a specific feature results in movement qualities.

Although endeavours were made to select musical excerpts with a relatively low level of dimensionality for the second observational study, this proved to be very difficult. True enough, the examples are simple and distinctly characterised by one or two features, but in most cases it was possible to include additional features that, it is reasonable to believe, interact with the most prominent features. One example of this is the Tchaikovsky excerpt. Here, I have initially described the music as mainly characterised by the dynamical shape (the *crescendo – decrescendo* contour). In addition to this we also have an *ascending – descending* pitch contour, a spread, dense texture, a full/warm timbre, a harmonic progression, and finally, a clear phrase (peak) structure. In the discussion of the dynamics feature I have suggested that changes in dynamics affect changes in activation level in the Tchaikovsky example. However, it seems clear that other features may also be contributing, for example the overall pitch contour and the more spread texture which results. This makes it difficult to assess whether one distinct feature prominently influences the activation level experienced.

7.6. Activation features in movement

In the review of features that are assumed to contribute to activation in music, I have already in the description of examples suggested features that in a similar manner affect the perception of global activation in body movements. These are listed in the following:

- *Density of events*, i.e. the frequency of onsets in time. Onsets is used here to mean changes of directions, accentuations in movement and the onsets of new chunks. High frequency leads to a higher level of activation.
- *Extension of movements*, i.e. extended, 'big' movements induce a higher level of activation.
- *Involvement of body parts*, i.e. movements that involve relatively large parts of the body evoke an increased sense of activation.
- *Tempo*, i.e. rapid movements result in a higher level of activation.
- *Force*, i.e. forceful and weighted movements result in a higher level of activation.
- *Articulation*, i.e. accentuated movements such as in abrupt changes of direction result in a higher activation level than when transitions are smooth.
- *Emphases*, i.e. a high frequency of weighted moments, such as goal-points, will contribute to a higher level of activation

This list is worked out on the basis of observing the videoed material collected from the dance-movements. They should be regarded as hypotheses about features that I believe contribute to the perception of overall activation in a gestural process. Again, the theoretical starting point is motor imagery; i.e. when we observe a sequence of movement, we imagine what it is like to perform the movements (the role of covertly simulating/emulating other people's movements while observing them was discussed in chapter 3, see Wilson & Knöblich, 2005). For example, if we see movements that increase in density we imagine what it might be like to increase the density of movements in our own body, and this imagery is supposed to evoke a sense of increased activation. Similarly, extended movements in space, movements that involve large parts of the body, quicker and forceful movements, as well as movements that are abruptly articulated, are all features that I assume influence the sense of an increased global activation level.

Clearly, as with the features that contribute to activation that were worked out for musical processes, these categories overlap. I shall come back to this point, but shall first briefly refer to examples from the collected video material to explain what I mean by each category.

7.6.1. Density of onsets

The category 'density of onsets' refers to the sense of how frequently 'new things' occur in the movements. 'New things' is used here to mean changes of direction (in any body part that is set in motion), changes in speed, subtle jerks that break up a smooth flow in the movement, emphasised moments (goal-points in a a peak-structured chunk), and onsets of new movement chunks. This means that 'onsets' refers both to the onset of the overall trajectory of a movement, as well as the onsets of what one might call more minor co-articulations subsumed in the main movement.

In one of the examples from the *Lento*, the dancer raises both arms in a slow and smooth movement, and she ends up with her hands meeting in a position above the head. There is a low density of onsets; i.e. within the time-span of approximately 8 seconds, 4 onsets are observable: (1) the segment is seen as one single chunk; (2) there is a subtle jerk as she interrupts the smooth flow by lifting her head; (3) there is a change in direction as she introduces a turn; and finally (4) there is weak emphasis when her hands meet above her head:

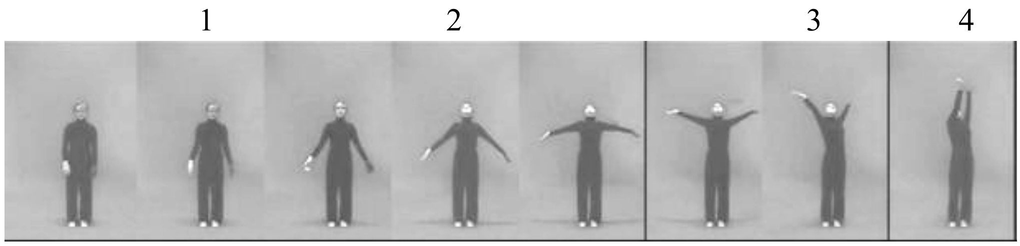


Figure 69. Raising the arms as a movement illustrates a sequence characterised by a low density of onsets.

7.6.2. Extension

Extension refers to the way the dancer uses and occupies the space around her body, so that if she stretches her arms out from her body, this is judged to be a high-level extension, and is accordingly thought to evoke a relatively high level of activation. In the example in figure 70 there is a change from a sequence with a low-level extension to a sequence with a high-level one.

Low extension — Low involvement
High extension — Higher involvement

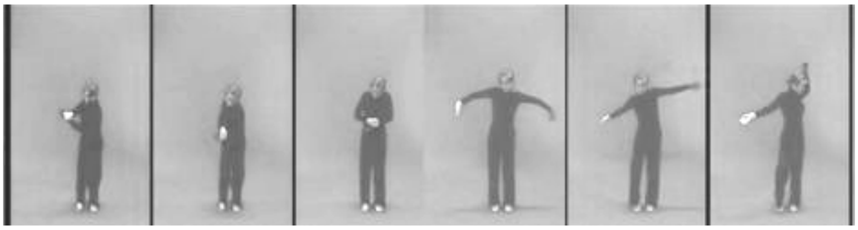


Figure 70. The push outwards exemplifies a change from a low to high extension and low to higher involvement.

7.6.3. Involvement

I would suggest that movements that involve large parts of the body (arms, legs, hips, torso), e.g. ones that spread fully throughout the body, may contribute to a higher level of activation than movements that are limited to a single arm-movement. In the example above there is in addition to the change in extension, also a change from the movement of the arms to a more spread movement that also includes the hips, torso and arms.

7.6.4. Speed

Speed refers to the displacement in time of one or more body parts. Not unexpectedly, the movements in the *Lento* variants are much slower than the ones in the *Prestissimo* variants. This observational category is closely related to density of onsets since a high density in most cases will require a higher speed.

7.6.5. Force

A forceful movement is used here to mean the movement of a body part seemingly performed with a strong and weighted muscular activity. It should be noted that force does not refer to exactly the same features as the *Weight* effort in Laban Movement Analysis, since *Weight* implies the connectedness of the movement, i.e. the way weight is transferred.

7.6.6. Articulation

Movement sequences that are performed smoothly, for example with gradual transitions between sequences and gradual changes in direction and speed will be termed smooth *articulations*, and these are assumed to be experienced with a relatively low level of activation.

7.6.7. Emphasis/goal-point

This final category captures moments of emphasis, e.g. goal-points that stand out from the gestural flow. Characterised as a more weighted onset, *emphasis* is closely connected to the density of onsets. I have included it as a distinct observational category since I assume that a higher density of *emphasised* onsets will make a qualitative difference, which suggests that such moments are more weighted by experience than 'any other' kind of onset. For example, in the *Prestissimo* variants there is in general a comparatively high frequency of emphases.

7.6.8. Overlaps between activation features

The way these individual categories have been explained points in the direction that there are considerable overlaps, as discussed earlier with respect to activation features in music. The overlaps between density and speed, as well as density and emphasis have already been mentioned. Furthermore, articulation will depend on both speed and force since a distinctly articulated movement will be experienced on the basis of an interrelated increase in speed and force. It is also reasonable to expect that extension in many cases will be linked to involvement, i.e. that a higher level of involvement ensues when the movements are more extended in space.

I have not attempted to work out distinct categories as first I do not think it is possible since I believe that we have to acknowledge that body movement, as with music, is a multidimensional phenomenon that is fundamentally characterised by the way aspects evolve in time as interdependent processes. Second, the categories are primarily understood from a descriptive perspective which means that they are not thought of as 'measurements' of single features that add up to a grand-total 'measurement' of global activation. Accordingly, they should be viewed as different angles from which to describe changes in activation; i.e. they are considered and described one at a time so that they can provide an approach through which to systematically observe activation.

7.6.9. Anvil as annotational tool

In the analytical procedure I shall use the Anvil annotational tool. I have worked out a specification file (xml-format) so that each observational category may be annotated at a time. For each category, different levels (low, moderate and high) are indicated by colours:

- Low – yellow
- Moderate – light red
- High – bright red

Yellow is thus associated with a low level of activation. Colour changes visualise for example how density changes according to the timeline. Although I would not assume that the features contribute to activation in a simple, summarizing manner, a darker shade in many of the features that occur within the same time-span might suggest that there is an increase or decrease in activation in this temporal segment.

The manually annotated features are then seen together using the machine-based QoM-analysis, as exemplified in figure 71:

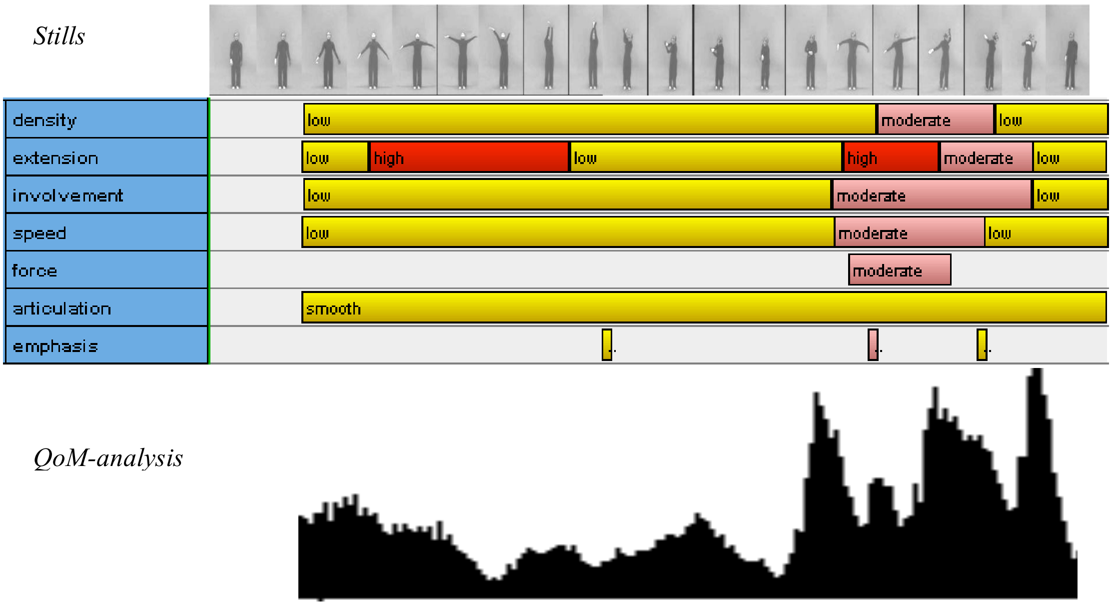


Figure 71. Annotation using Anvil of activation features combined with a QoM-analysis

Viewed together, the manually annotated analysis and the computer-based QoM-analysis describe how activation changes over time, and how different aspects of movement may contribute to these changes.

So far I have mainly discussed the connection between individual features and the way each of them may contribute to global activation in music and body movement respectively. But the observational categories used in the air-piano study imply that certain features of music, such as loudness, directly correspond to defined features of air-playing such as size and speed. Considering the openness discussed above of the free dance-movement task, I would preliminarily assume that a change in one specific activation feature in music (e.g. an increase in density of onsets in music) does not necessarily and consistently correspond to a similar change in one specific activation feature in movement (e.g. an increase in density of onsets in movement). This means that I would hypothesise that the correspondence is flexible in the sense that increased density in the music may evoke a higher level of activation, but that this increased activation may be reflected in other activation features in the movement such as extensions or articulations.

On the other hand, it does seem that some of the music features may have a more direct link to specific movement qualities. The examples from the second observational study suggest that the dancer responds consistently to differences in articulation.

A similar link may be found for the feature of loudness since it appears that music played loudly tends to be reflected in movements that are extended in space, and/or that involve large parts of the body. This seems to be in accordance with the findings from the sound-tracings study. Both sound-tracers and interraters commented that the sonic examples that were characterised by a dynamic swell could not be properly visualised using the restricted format of the tablet, and that they would have preferred to visualise the swell using "something" that occupied more space. I would interpret this to mean that loudness may be related to the observational categories for movement activation that I shall later refer to as extension and involvement. I think this connection was demonstrated by the Tchaikovsky example discussed earlier in this chapter (music from the opening of the second movement of the 5th Symphony, (see audio-visual example *avex18* on CD). But again, considering the multidimensionality, it is difficult to evaluate whether it is the loudness or the spread-out texture that makes up the link to movement qualities.

7.6.10. Activation features summarised

This last issue, the flexible correspondence from a specific activation feature to a global activation contour vs a more direct link between specific activation features, seems to be closely connected to the recurrent theme of this thesis, i.e. the *flexible* vs the *non-arbitrary* nature of music and movement correspondences.

To sum up the analysis of activation in this section, the following points are important:

- Two sets of activation features, one for music and one for movement, have been worked out and explained.
- The individual activation features are thought to contribute to global activation in music and movement respectively, though not in a simple, summarizing manner.
- The activation features overlap and are partly interdependent on each other which reflects the multidimensional character of both musical and gestural processes.
- The two sets of activation features will be used to annotate the audio-visual material to describe the way activation changes over time, and furthermore, how different features contribute to these changes in the variants analysed.
- The two sets will also form a basis on which to assess whether specific features in music correspond more directly to specific features in movement.

7.7. Basic effort actions and dynamics

As discussed in chapter 4, activation is understood as a blend of the dynamical and kinematical aspects of movement, for example shadings of *articulation* are captured by what we see, by the rate of change in direction and speed (kinematics), as well as in light of the way these changes evoke an imagery of the way forces cause and constrain the movement observed (dynamics).

Additionally, I shall attempt to describe aspects of music and movement that I understand as primarily dynamical. For example, at certain moments in the *Lento* there is a quality to the timbre that appears increasingly to be pressing or squeezing. In the analysis of movement variants I shall discuss whether correspondence is influenced by the way the dancer captures and expresses this quality.

The theoretical point of departure for this sort of description is Laban Movement Analysis. As outlined in chapter 4, Laban grounded his effort theory on the notion that the study of movement requires the process be examined on multiple levels, i.e. that movement is understood as combinations of effort qualities according to the motion factors of *Weight*, *Time*, *Space* and *Flow*. Hence, he proposed eight *basic effort actions*, which are in fact variants of the two *basic effort actions* of “thrusting” and “floating” (Laban, 1971). In table 20, they are listed and it is explained how each of them are characterised by effort qualities. The table shows how Laban suggests that dabbing, pressing and slashing are derived from thrusting by replacing one effort quality (underlined). Likewise, wringing, flicking and gliding are related to floating. The table also includes further derivatives of the basic effort actions, i.e. movements that are characterised by similar effort combinations[107].

Basic effort action	*Weight*	*Time*	*Space*	Derivatives
Thrusting	Firm	Sudden	Direct	Shove, punch, poke
Dabbing	<u>Gentle</u>	Sudden	Direct	Pat, tap, shake
Pressing	Firm	<u>Sustained</u>	Direct	Crush, cut, squeeze
Slashing	Firm	Sudden	<u>Indirect</u>	Beat, throw, whip
Floating	Gentle	Sustained	Indirect	Strew, stir, stroke
Wringing	<u>Firm</u>	Sustained	Indirect	Pull, pluck, stretch
Flicking	Gentle	<u>Sudden</u>	Indirect	Flip, flap, jerk
Gliding	Gentle	Sustained	<u>Direct</u>	Smooth, smear, smudge

Table 20. The basic effort actions as combinations of *Weight, Time* and *Space*.

In the analytical work later in this chapter I shall explore how musical and gestural processes may be described by metaphors, such as those suggested in the table above. This means that I shall elaborate on this part of the analysis on the understanding that shadings of movement qualities emerge from the way aspects combine and interact, and that some of these movement actions are more closely related than others. For example, in the *Lento variants* the movements are typically performed with a sustained *Time* effort combined with either a firm or gentle *Weight* effort, so that they may be characterised as either pressing or floating.

On the Anvil annotational-board, this aspect of movement is referred to with the term *dynamics.* This feature will also be addressed in an analysis of the musical excerpts; for example the *Prestissimo* is described by a change from a ‘light, nervous, jittery jumpiness’ to a ‘heavy, slow, tenacious flow’. The term dynamics will be used interchangeably with the expressions *movement character* (used in the movement analysis) and *dynamical shaping* (used with reference to changes in the music). I have considered the term *flow* since the feature we are talking about has something to do with the way the movements and music appear to flow in time: lightly, heavily, easily, slowly, and so forth. But I have avoided the term since it may interfere with Laban’s term referring to the *Flow* effort.

[107] The reader may have observed that the *Flow* factor has not been included. In Laban Movement Analysis *Flow* is commonly treated as an additional *effort* element that further colours a movement, e.g. that thrusting may be performed with a free or a bound *Flow* effort.

7.7.1. *Effort* vs *activation*

It may be argued that activation and effort overlap, so that it is not necessary to consider both approaches. For example, *Time* (effort) and *speed* (activation feature) are apparently very similar. The same applies to *Weight* (effort) and *force* (activation feature). Although while exploring the movement variants I have also addressed effort qualities and annotated changes in motion factors in the Anvil annotational board, I shall not systematically include effort analysis in the proceeding analysis. The way of thinking about movements as multi-dimensional that underlies Laban Motion Analysis will of course be included as outlined above. Additionally, I shall occasionally include effort analysis when I think that specific effort qualities describe movement sequences more appropriately. Although activation and effort certainly overlap, they may also be seen as complementary aspects. As reviewed in chapter 4, effort is understood as a quality that borders on intentionality or the mental attitude of a movement. A series of quick movements performed at the same overall speed by two dancers may turn out differently, i.e. one may be experienced with a 'rushed' intentionality, the other as 'calmer'. An effort analysis might reveal that in the second case the gestural flow is characterised by a smoother fluctuation between firm and gentle *Weight*, so that the movements appear to be more grounded and connected. This means that the exploration of musical and gestural processes in terms of metaphors, as suggested above, introduces an additional aspect of meaning, suggesting that *activation* is here understood to be a more neutral, physical aspect.

7.8. Event features: chunking, shaping, phrasing and synch-points

So far in this section I have discussed how the concepts of activation and effort will be understood and implemented within the context of the following concrete analysis. This leads to the final question of how these aspects are included in temporal processes in terms of event features such as shaping, phrasing and patterning (see chapter 5). This means that music and movement correspondences will also be analysed with respect to the way the two streams appear as simultaneous events characterised by a beginning/ending (chunking/segmentation), kinematical and dynamical shapes, and the way emphasised moments in music and movement are synchronised (synch-points). The following questions will be addressed:

- Do the overall activation contours of music and movement correspond to each other within a temporal window?
- Do they have a similar phrasing structure in terms of synchronised moments of emphasis?
- Are music and movement perceived with simultaneous beginnings and endings in terms of chunks?

Features related to chunking will be systematically observed and annotated in the Anvil annotational chart. The example in figure 72 illustrates how observation of chunking and shaping in the gestural variant has been broken up into three tracks: chunks (the immediate observation of beginnings and endings), kinematics (beginnings and endings of trajectories) and dynamics (characterisation of dynamical features, e.g. pressing, shoving, pushing, etc.).

Figure 72. The movement sequences are observed and then annotated in Anvil in terms of chunks, kinematics and dynamics.

As discussed in chapter 5, the perceptual process of chunking is closely connected to the way we perceive the peak structure of an event. In an audio-visual compound, a simultaneous accentuation in the music and movement may emerge as a synch point, i.e. a goal-point of an audio-visual chunk. In this regard the emergence of synch points will be considered both in terms of the simultaneity of musical and gestural accentuations, and in the way accentuations are a part of the temporal context. This means that synch points are not only a matter of timing; they are also a matter of how an accentuation is prepared, i.e. whether the musical and gestural flow push towards the accentuation in a similar manner so that music and movement together afford the perception of a clear audio-visual peak structure. This interplay of timing and dynamical aspects in the emergence of synch points will be observed and analysed.

7.9. Audio-visual analysis: systematically analysing variants and alternating between an alone and combined condition

The analyses in this chapter of music-movement correspondences will attempt to make use of the descriptive potential that lies in the study of differences. This is the reason for including twelve different variants in the case study. The material consists of variants that demonstrate different nuances of correspondences. In this lies the descriptive potential: by comparing variants in terms of the different degrees or the nature of correspondence there is a potential for discussing how qualities in music and movement contribute to audio-visual correspondence.

A further challenge for audio-visual analysis is the role of *interaction effects* due to multisensory integration. This was discussed in chapter 3 and is briefly summarised as follows:

- Perceived correspondence between an auditory and a visual sensory stream emerges on the basis of some kind of similarity between the two streams.
- Perceived correspondence may also to some degree emerge as a result of two streams occurring at the same location at the same time, i.e. it seems that the perceptual apparatus tends to fuse ‘things’ together into one audio-visual event.
- The auditory and the visual stream are capable of modifying each other reciprocally. The modality-appropriateness hypothesis suggests that audition overrides vision in tasks characterised by temporal features, whereas vision overrides audition in spatial tasks.

Appreciating such effects as a prominent feature of audio-visual compounds implies that it is advisable to approach the analysis under three different conditions, i.e. music alone, movements alone and music and movement combined. The reason for this is that we expect music and movement respectively to appear differently in the alone condition compared to the combined condition. In other words, I would suggest that studying correspondences solely on the basis of the way music and movement are heard and seen separately, i.e. a comparison of analogous qualities, would not give the full picture. Looking closely at what happens when the streams come together is also required, i.e. how these qualities in music and movement respond to each other, whether they are drawn to each other, or whether they are seen-heard as separate streams.

Chion has proposed a masking method which means that the audio-visual analysis is approached by alternating between the three conditions mentioned above (Chion, 1994). The output is descriptions of the different appearances of music and movement, and comparing the alone versus the combined conditions is thought to reveal an insight into the audio-visual combination as a unique form of articulation, i.e. the audio-visual form reveals itself in the differences between the way music and movement respectively appear in the alone versus the combined condition.

Similarly, the question of differences between the alone and the combined condition is central to the experimental design in studies that have examined interaction effects, e.g. perceived intensity and segmentation (Krumhansl & Schenk, 1997; Sirius & Clarke, 1994; Vines et al., 2005).

The methodological key word is *differences* since audio-visual effects are discussed on the basis of observed differences. In my view, studying differences has a much more far-reaching potential than this, which is to say that the differences do not only say something about interaction effects, but also something about specific features in the music and movement as they are heard and seen before they are combined. The differences inform one about both what music affords before it is combined with something visual, as well as what the music offers in combination with a certain sequence of body movements. This means that the analytical potential goes both ways, from the alone condition to the combined condition and back again. And this leads to the methodological consideration that thorough analyses of music and movement in either of the conditions, the alone as well as the combined, are essential for an understanding audio-visual correspondences. Although it has been proposed that interaction effects, among them music-movement correspondences, should be understood as an emergent quality of perception, I think that correspondences, or the potential for correspondences to perceptually emerge, are constrained by what is afforded ‘in’ the music and the movement before combination.

Furthermore, *difference* is a methodological keyword when employed as a descriptive strategy. Music and movement, as well as music-movement combinations, are very difficult to describe because of their transitory nature. We may try to describe a musical performance and the way we experienced it in different ways, but very often we start describing the performance with reference to other performances, i.e. the musical experience reveals itself in the way it differs from other experiences.

7.9.1. Degree and quality of correspondence

One part of the analytical procedure will be to judge and describe the degree and quality of correspondences in the twelve variants. Different terms and expressions reflect variations in the way music and movement relate to each other. In some cases, the relationship appears to be very close, as though music and movement were 'merging', so that 'convergence' or 'melting together' may be the proper descriptive terms. This situation resembles the everyday situation in which a sonic and a visual event are integrated because they emanate from the same source; in fact they are tied together by being two articulations of the same event. This is what happens in the post-synchronisation of film – i.e. the sound of steps is captured by seeing the steps. In an everyday situation, the audio-visual link between sound and seeing the steps is real; in the post-synchronised situation the link is virtual. However, the experiential quality is similar: in both cases image and sound are intimately linked as though they were emanating from the same source.

One example that demonstrates this tight kind of correspondence is found in one of the variants performed by the dancer B. In the final movement of this variant she pushes both arms outwards and the dynamics of the movement seem to merge with the phrasing of the music.

In other cases, the correspondence appears to have a looser quality. One example of this is the opening of one of the variants performed by dancer A. The musical component is characterised by a sustained timbre played on woodwind, and the corresponding movements are two successive smoothly and slowly performed arm waves. Both the music and movement have a sustained, stretched quality, but considering the higher level of activity, the movements appear to float on top of the music, as though the movements added an extra textural layer to the music.

The variants also demonstrate that they may be described both in terms of differences in overall correspondence, as well as with respect to the way the correspondences differ within one single variant. The part of the analysis that concentrates on the combined condition will in various ways attempt to describe all aspects of correspondence and synchronicity. To this end I shall try to find terms/metaphors that most precisely reflect my own interpretation of the music-movement relationship. I will use the Anvil annotational board for the degree of correspondence (see figure 73). This procedure enables an initial visualisation of the way correspondence varies along the timeline of a variant. The example is an annotation of correspondences in the variant performed by dancer A, referred to above. The first track of the annotation shows that I judge the degree of correspondence to vary between moderate and high. The next track in this example identifies two points of synchronisation, i.e. moments when accentuations in both the music and movement are perceived as synchronised. The first of these is judged to be weaker than the second one.

Chion proposes that in addition to judgement of the degree of correspondence and the identification of synch-points, the audio-visual analysis at this stage should also address the questions: *"What do I see of what I hear?"* and *"What do I hear of what I see"* (Chion, 1994; p 192). Additionally, and as an elaboration on this kind of thinking, I would propose that a related question should be asked: *"What do I see that*

I do not hear". As seen in the example in figure 73, these questions have been implemented in the annotation board. Descriptions of this kind serve the purpose of providing an initial judgement of the way movement and music work together. In this case they suggest that the arm waves are to some degree perceived as different from the musical flow in the first segment of the variant. Furthermore, the annotation points to another segment with tighter correspondence, suggesting that a lift in the movements matches a stretching quality in the sonic component.

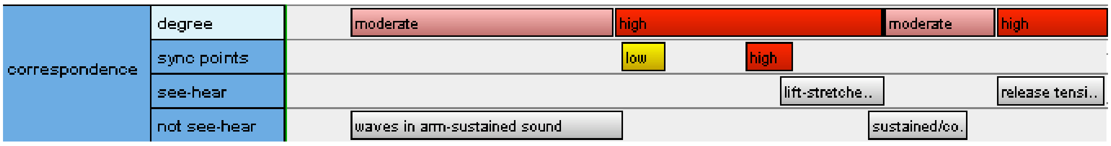


Figure 73. Excerpt from Anvil annotational board that is used to describe variations in correspondence, the distribution of synch points, as well as the way variations in correspondence may be explained with reference to similarities ('see-hear') and differences ('not see-hear') in music and movement.

In parallel to the descriptions and analyses of correspondences, I shall discuss features in music and movement as they appear in the alone condition. During this process the analysis will concentrate on the qualities that I assume play a part in perceived correspondence, with the initial correspondence analysis as reference point. However, I shall also attempt to approach the analysis quite open-mindedly and from different perspectives so that subtle shadings are also considered.

The main emphasis of the audio-visual analysis is to describe all aspects of correspondence in the selected material. This includes describing how features of music and movement merge in different ways; how they occasionally meet in tight synchrony and in other instances how they separate into two streams. Since this objective requires thorough investigations of music and movement in both the alone as well as in the combined condition the two conditions are understood as equal reference points.

Thus, by alternating between analyses of music and movement as they appear in the alone and the combined conditions, the aim is to describe the different ways in which music and movement correspond. For example, in light of the analysis of the degree and quality of correspondence ('strong' vs 'weak'; 'amalgamation'/'fusion' vs 'separate' streams), I shall, in each case, discuss the way specific features of music and movement respectively may contribute to correspondence.

Chapter 8. Analyses of music-movement correspondences in the *Lento* and *Prestissimo* variants

Building on the analytical procedures and terms worked out in the previous chapter, I shall in the following discuss similarities and differences between music and movements in audio-visual variants based on two different musical excerpts from Ligeti's *Ten Pieces for Wind Quintet*, referred to from now on as the *Lento* and *Prestissimo.* The two musical excerpts differ markedly, the first one being slow and sustained, the other characterised by quick, leaping figures. I shall first describe the musical excerpts in detail, and then describe general similarities between the gestural variants. This is followed by a more detailed analysis of a few selected audio-visual variants.

8.1. Analysis of the *Lento* excerpt

This excerpt is taken from the opening 20 seconds of the *Lento* movement from Ligeti's *Ten Pieces for Wind Quintet*[108]. It is characterised by its slow, gradually evolving quality. The sequence opens with a sustained, mildly dissonant timbre, composed of a cluster of three tones played on the flute (contralto), oboe and bassoon (see figure 74). Out of this homophonic texture a tone (flute, contralto) gradually grows and comes to the foreground, thus initiating a short sequence that may be denoted as having light, 'rippling' figures. The higher density of tone onsets at this stage suggests these figures. Finally, a short melodic line appears. It consists of two tones played on the oboe. The last of these tones, a C#4 [109] is resolved by a D4 in the flute, so that these three tones constitute a melodic line. The excerpt fades out to a more dissonant polyphonic texture.

[108] Ligeti, G. (1998). Ten Pieces for Wind Quintet. On *György Ligeti Edition Chamber music* [CD]: Sony Music Entertainment Inc. Performed by London Winds

[109] Midi notation

‘Sustained timbre’ ‘rippling figures‘ ‘three tone melodic line’

Figure 74. Excerpt from score[110]. The parts are notated in C

This initial description provides an overview of the way tonal elements are distributed in time. The next step is to attempt to characterise change, first by using metaphors to characterise impressions from the music and the way these are related to tonal elements, and second, to analyse the music in terms of its phrasing pattern.

The opening timbre has a light, gentle and transparent quality. This changes very gradually and subtly as the flute tone crescendos, thus causing a more insistent and brighter timbre. The emerging timbral colour is more detained; it is as though the tonal flow is squeezed into a narrower space. The tension is released as the B3 (flute) is introduced in the second bar. At this moment the flow opens up, as though a deep breath were being taken. Also at this moment, it appears that the music gains weight, i.e. in a brief moment of emphasis the effort (Laban) may be characterised as a firm *Weight*, as opposed to the preceding gentle *Weight.* The change from the first to the second bar can be characterised by imagining a body that, by the end of bar 1, is at the end of expiration and is pushing out the air, and then opens, or releases the tension of the body to take in a new breath.

The following sequence, referred to above as ‘rippling’ figures, produces an impression of expectation or waiting. Then, as the melodic line is introduced a distinct dynamic shape emerges. The phrasing of the figure resembles something opening up by sustainedly pulling, e.g. pulling the arms out from the body in a stretching manner, followed by a similarly sustained contraction. The dynamic shape is also characterised by an emphasis; the phrasing pushes gently forwards to a goal-point at

[110] Ligeti, G. (1969). Zehn stücke für bläserquintett (stimmen/parts) [Sheet music]: Schott. The score is notated in C.

the C#4 (oboe). The emphasis is roundly, smoothly articulated, characterised by a firm, shoving push. At this moment the music gains firmer *Weight* (Laban effort).

To sum up these descriptions, the music affords a number of changes in features, listed in table 21 with the oboe and flute staves as a coarse timeline reference:

			
Metaphor	(open) narrow	open	opening - closing
	stationary subtly pushing	awaiting	pushing - withdrawing
	gentle squeezing	firm	firm gentle
			pulling up - contracting, clenching
Intensity	low increasing	decreasing	increasing-decreasing
Texture	homophonic, dense	polyphonic, more open	
Timbre	light, porous brighter		brighter
Loudness	pp cresc.		cresc. - decresc.
Density onsets	low	moderate	moderate
Articulation	smooth	smooth	smooth

Table 21. The *Lento* excerpt is characterised by changes in a number of features.

The descriptions of change in table 21 suggest an overall phrase structure that may be characterised as being two successive waves, the first one long, stretched and vague, and the second more apparent and characterised by a more rounded dynamic shape. In the drawing in figure 75 the curve is meant to visualise my own interpretation of the phrasing, i.e. a rising curve indicates that the music is pushing forwards, and arrows indicate the timing of emphases (the dashed style of the first arrow suggesting a less weighted emphasis). The curve may also be interpreted as a visualisation of changes in intensity/brightness:

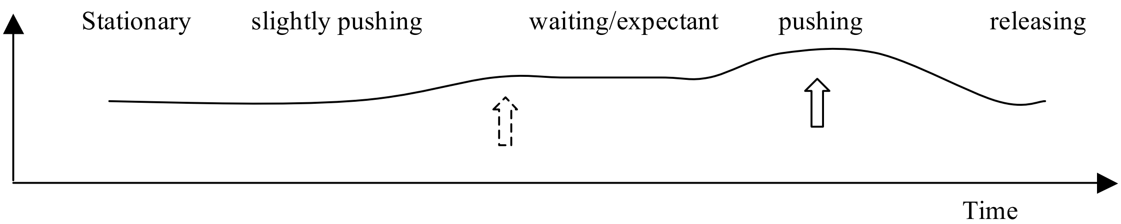


Figure 75. The curve visualises the phrasing in the *Lento* excerpt

The phrasing outlined above may be further examined in analyses of the sound signal. The waveform irepresentation n figure 76 visualises the two waves, i.e. that they are partly a result of two successive 'increase/decrease' patterns in *loudness*.

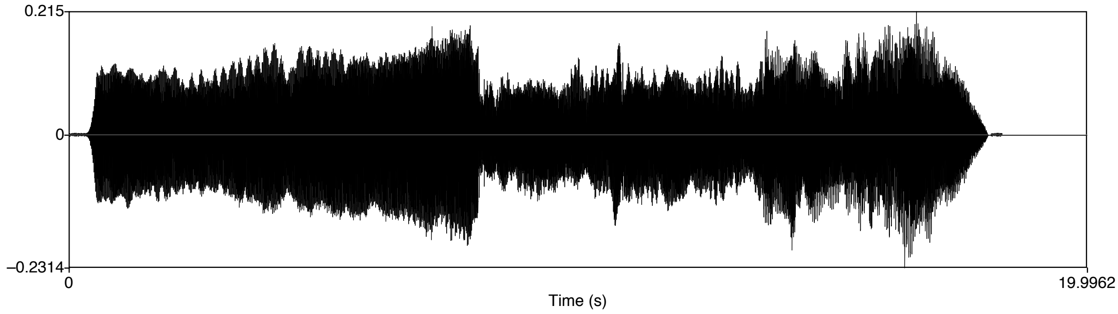


Figure 76. The waveform visualises two successive 'increase/decrease' patterns in loudness.

Furthermore, in the spectrogram in 77 I have superimposed the visualised interpretation of phrasing. Scattered grey areas in the higher frequency range suggest a slight increase in brightness just before the first moment of emphasis (indicated by an arrow). We may also observe a more marked increase in brightness after the second emphasis represented by a number of strong frequency components in the higher ranger. The spectrogram also suggests that the second emphasis is stronger than the first. In sum the spectrogram visualises that the phrase structure is also affected by two successive 'increase/decrease' patterns in brightness.

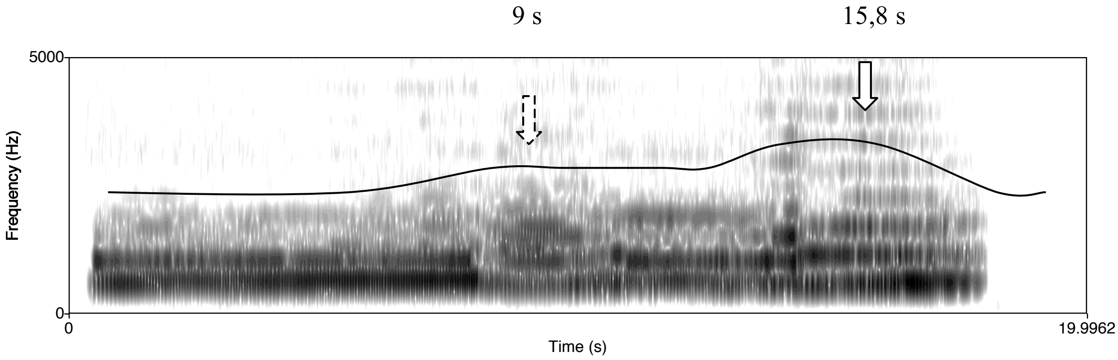


Figure 77. The spectrogram suggests two successive 'increase/decrease' patterns in brightness.

8.2. Analysis of *Prestissimo* excerpt

The musical component in this example is the first 15 seconds of the fourth movement *Prestissimo leggiero e virtuoso* from Ligeti's *Ten Pieces for Wind Quintet*. Initially, the excerpt may be described as one long phrase which starts with activity at a relative high pitch register. This activity has the character of something light and nervous that jumps around in an unpredictable manner. Another image that comes to the mind is the image of someone trying again and again to grab something very light and 'fleeting'/'flighty' such as floating feathers that escape by seemingly wriggling away from these hands or that are blown away by the air. A calmer character towards the end of the excerpt succeeds this first sequence and alludes to quick, abrupt and jittery movements. Thus the quickness of the opening is contrasted with a slower and softer character perhaps resembling a liquid flowing slowly, softly and heavily.

To address the musical elements more directly, the first part consists of interwoven figures played mainly by the flute and clarinet[111]. The flute voice appears to be the primary foreground, but occasionally the clarinet takes over. The pitch space is spread out within an octave, and sometimes more, in the first bars. There is a high density of onsets and the figures are characterised by their leaps between pitch registers. The tones are played lightly with an airy quality, some of them more distinctly articulated.

Figure 78. Excerpt from *Prestissimo* score (notated in C).

I have tried asking myself whether any kind of rhythm is emerging. Certainly, there are contrasts in the way tones are articulated and emphasised, but it is difficult to pin down any kind of temporal pattern in these contrasts from listening alone. The more emphasised moments appear to be very transitory; they are not heard as moments of gravity, and they do not occur periodically, i.e. a clear pattern is not established. The musical figures are instead heard as short, ‘choppy’ waves, characterised by their sudden turns and abrupt changes, which do not find rest or let the weight fully into a moment of gravity. The composer indicates this overall non-periodic and non-metrical rhythmic character in his performance instructions: “Apart from the indicated accents, always play very evenly and without accentuation, so that the subdivision into bars does not become perceptible”[112]. This feature is further brought out by the timing, phrasing and articulation of the melodic figures. The quarter note alternates between being subdivided into three and into four, and the accents are placed, apparently non-systematically, on different tones within the triplets and quadruplets respectively.

In the concluding part, which corresponds to bars 7 - 9 in the score, the musical characteristics mentioned above are changed. We reach a lower pitch, the pitch space is restricted to a major third, and the onset density is considerably reduced. The tones

[111] Ligeti, G. (1969). Zehn Stücke für Bläserquintett (Stimmen/parts) [Sheet music]: Schott.
[112] Ligeti, G. (1969). Zehn Stücke für Bläserquintett (Stimmen/parts) [Sheet music]: Schott.

are more smoothly articulated, and the timbre is less bright. These last bars are played by the clarinet and the bassoon, which means that the texture has been thinned out. Furthermore, there is a change from a dense polyphonic texture to one more homophonic and characterised by its slowness and tenacity.

To sum up, the music is characterised by its overall change from a high to a low activation contour; it is as though a burst of energy and tension is poured into the music at the beginning. This energy gradually fades as though the music were getting 'out of breath' towards the end. The change is characterised by going from a light, nervous, jittery jumpiness to something heavy, tenacious, rubbery and slowly flowing. This overall change in character may be connected to the changes in musical features that I have discussed, i.e. the gradual changes in density (high to low), pitch space (sudden leaps to concentrated), articulation (sharp to soft), and timbre (bright to dull), as well as the overall pitch contour going from high to middle register. These changes may be illustrated graphically by the same shape/curve which also corresponds to the overall activation contour.

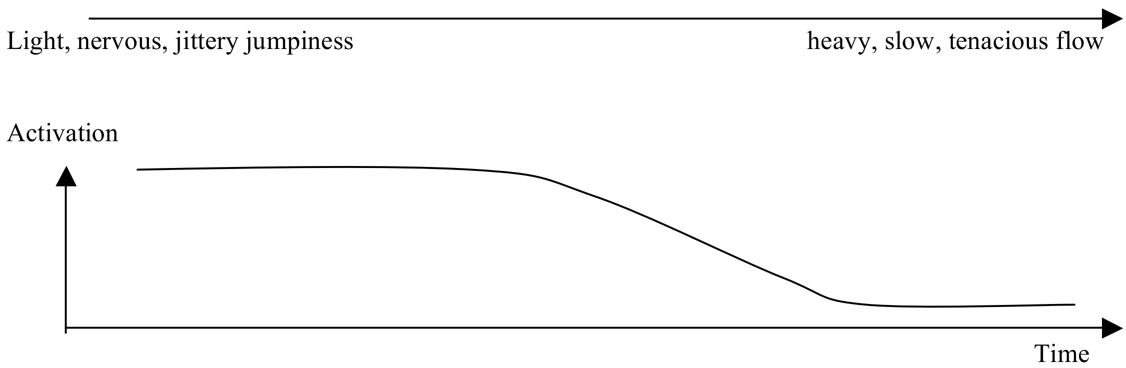


Figure 79. The curve schematically suggests the change from a high level to a lower level of activation.

The changes that characterises the excerpt are further illustrated by analyses of the signal. First, the waveform representation shows how the amplitude of the signal changes as a function of time, i.e. from high to lower.

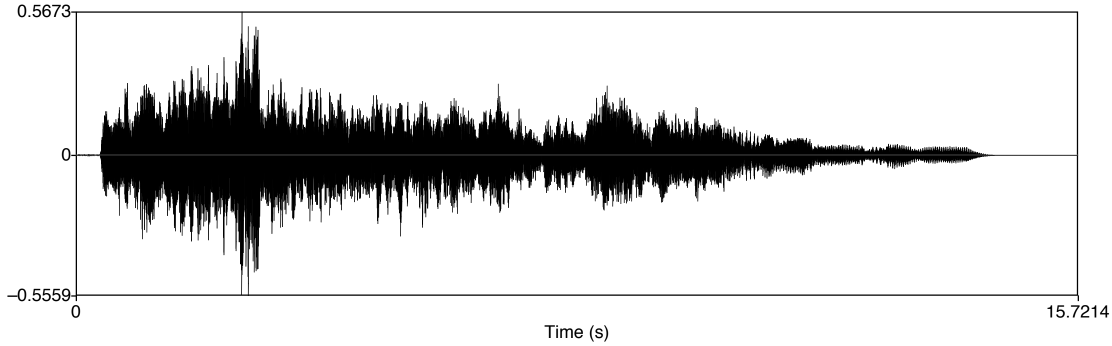


Figure 80. The waveform representation indicates a change in loudness from high to lower.

Second, the spectrum shows stronger and more distinct frequency components at the beginning and more blurred grey tones towards the end, which points to changes in timbre from bright to less insistent.

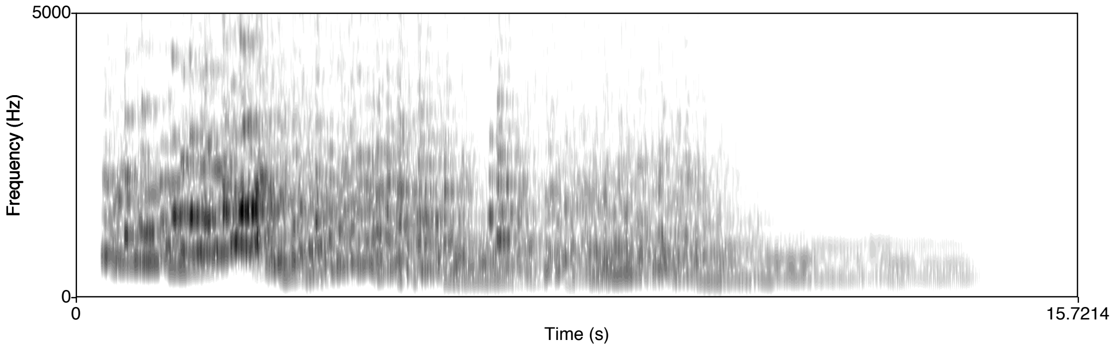


Figure 81. The spectrogram indicates a change from bright to less bright towards the end.

Finally, the pitch extraction (all instruments included) visualises the change from a spread-out pitch range to a more restricted range:

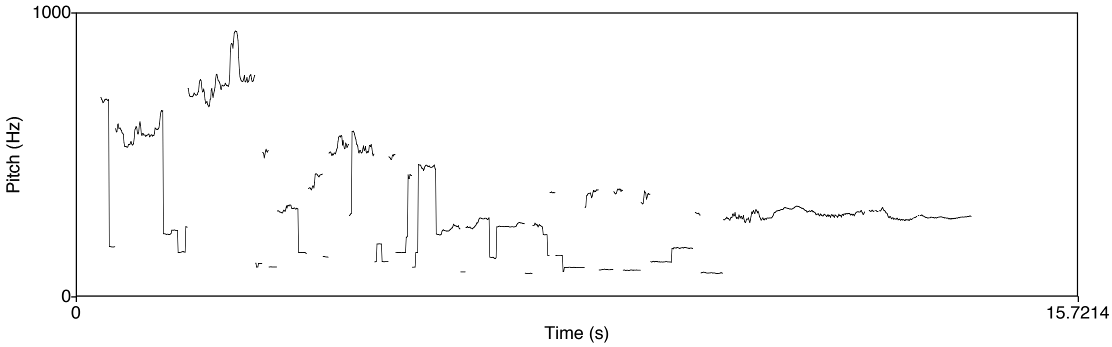


Figure 82. The pitch extraction visualises the change in pitch range.

8.3. Initial analysis of correspondences between music and movement variants

The tables on pages 214 to 217 show the Anvil annotations and QoM-analyses of twelve dance variants, six *Lento* and six *Prestissimo* variants performed by two female dancers (dancer A and dancer B). Tables 22 and 24 display the analyses of the *Lento* and *Prestissimo variants* performed by dancer A[113]. Tables 23 and 25 display the analyses of the *Lento* and *Prestissimo variants* performed by dancer B[114]. The first row in the tables shows the qualitative analysis of chunking and how the chunks are distributed along the timeline annotated in *Anvil.* Time runs from left to right and vertical bars show the beginnings/endings of chunks so that one box represents one chunk. The analysis was performed on the basis of repeated observations of the movement performance in the alone condition while addressing the questions: "When do I perceive a new chunk starting, and when do I perceive the chunk ending and a new chunk beginning?"

The question is: which criteria do we use when we judge which elements in a movement sequence belong to the same chunk? Earlier in this chapter it was noted that limitations in the short-term memory constrain chunking in terms of the duration

[113] The variants are available as mpeg4-files on the CD
[114] In the reference A-1.1, the letter *A* refers to the dancer and *1.1* refers to the first interpretation of the first musical excerpt

of a chunk. The question of chunking criteria was also discussed in chapter 5. With regard to music, pitch features were mentioned as one aspect that contributes to chunking, e.g. that tonal elements that proceed in a stepwise manner facilitate these elements being 'glued' together into one chunk, as in a melody line (see Bregman, 1990). This criterion is illustrated in the *Lento* excerpt in the stepwise line towards the end of the excerpt. Within the same time window, the same passage may be viewed in terms of a *kinematical* and a *dynamical* shape, i.e. *kinematics* understood as the 'rising' pitch contour and *dynamics* understood as the 'increase/decrease' pattern in intensity (referred to as a 'wave' with a 'rounded dynamic shape'). Finally, the performance of the *Lento* excerpt suggests a *prefix – goal-point – suffix* phrase structure (previously also referred to as a *peak structure*) in which the goal-point is assumed to play an important role in chunking.

The same features (except for pitch relations) are relevant in the chunking of movements. In general, recognising global features, referred to as *shapes* or *contours*, includes perceiving the (approximate) timing of the beginning and ending of a shape, hence the beginning and ending of a chunk. The criteria for chunking in movements are summarised as follows:

- *Dynamical shapes*, i.e. does the movement afford a clear contour of changes in intensity/activation? To illustrate this, the series of stills in figure 83 shows a movement (from one of the *Lento* variants, B-1.1) in which the dancer pushes her arms outwards and pulls them back again, the movement suggesting an 'increase-release' of energy.

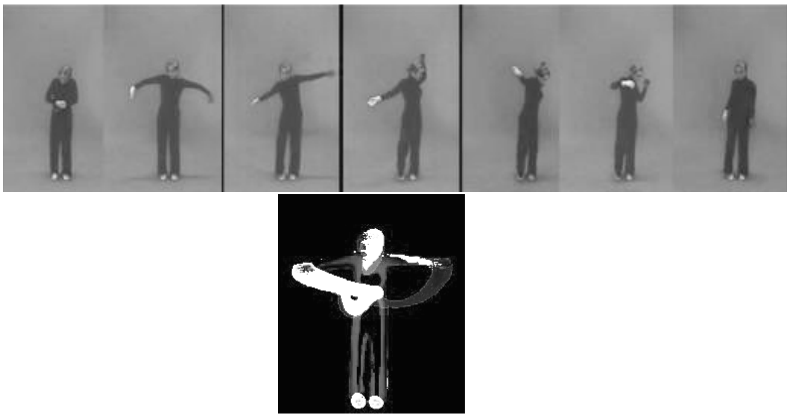


Figure 83. The still series and the motion-history image show the outwards push from one of the *Lento* variants.

- In this sequence the push outwards is seen as an emphasis so that this point of accentuation appears as a goal-point. Thus, the example also illustrates the notion of *prefix – goal-point – suffix* and the way this kind of structure contributes to chunking.
- *Kinematical shapes*, i.e. does the movement afford clear contours in terms of trajectories in space? This is illustrated in another of the *Lento* variants (A-1.1, see series of stills, figure 84) in which the dancer performs two successive, similar kinematical shapes that are characterised by the trajectory of the right arm. She lifts the arm away from the body and back again in a slow, wave-like movement.

This 'excursion' of the arm is seen as one coherent chunk.

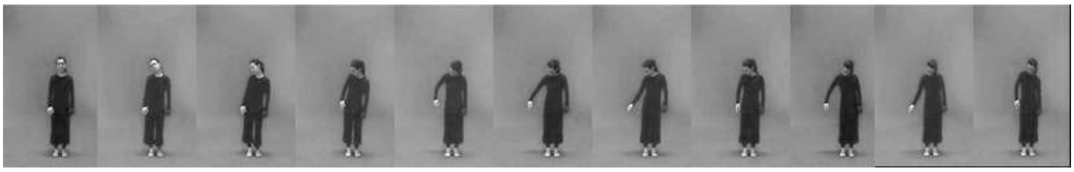


Figure 84. The still series shows the excursion of the arm as one coherent chunk.

- The example also exemplifies the role of patterns in chunking. Here, two similar movements succeed each other; the first 'wave' is repeated so that a pattern emerges. The repetition offers the possibility of recognition, which in turn provides a clue for making a judgement about where "something" ends, and at what point "something new" begins.
- Additionally, to the extent that the movement alludes to symbolisation, e.g. that we actually associate the movement with a wave, this may further contribute to solidifying the movement sequence as one chunk.

As mentioned previously, the dancers in our observational study demonstrate a richer and more precise movement repertoire than novice performers. This should be viewed in light of the strategy they appear to use when faced with a task such as in the videos we made. Through their training as dancers they have established a rich pool of well-rehearsed movement variants and figures. The videos show that they typically use a few movement patterns (e.g. the arm waves referred to above), which are repeated, varied and transformed. Using this approach, the movements become precise in the sense that the emerging patterns facilitate chunking.

One final note should be added about how precisely the timing of the beginnings and endings of a chunk is perceived. In chapter 5 it was commented that in many cases it appears that two successive chunks are perceived with a certain overlap, i.e. that in many cases it appears that 'something new' has already started, although the 'previous something' has not quite ended. This issue of overlapping should be borne in mind during the process of analysing since:

- first, in some cases in which the transition between movements is smooth, it may be difficult to judge decisively whether one sequence consists of one or two chunks (e.g. the two arm waves discussed above may be seen as one coherent movement unit);
- second, what one might term the box format used in Anvil annotation, suggesting that beginnings and endings occur at precise moments, does not fully reflect the often blurred boundary between successive chunks. In other words, the timing of beginnings and endings in the annotation should not be taken too literally.

As with the chunking feature, the activation features are annotated on the basis of repeated observations, and one feature at a time is addressed, e.g. "How would I characterise the onset density, and how does onset density change over time?" The yellow boxes indicate a low level of activation, whereas the light red and dark red boxes indicate relatively higher levels of activation (moderate and high).

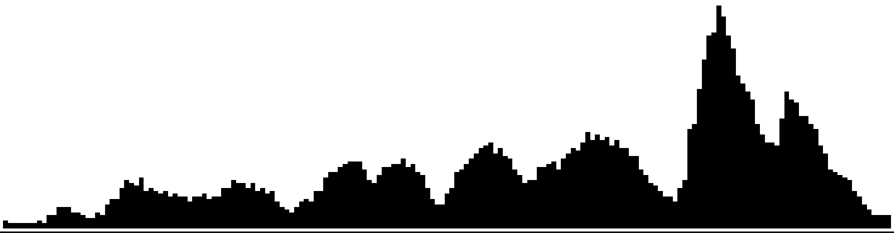
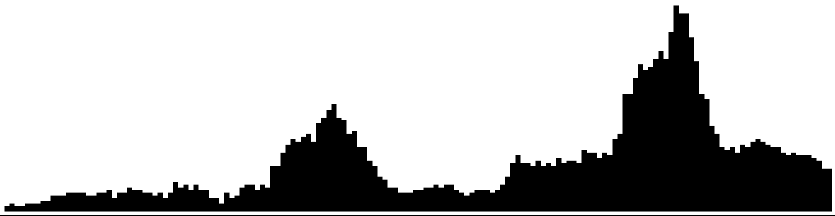
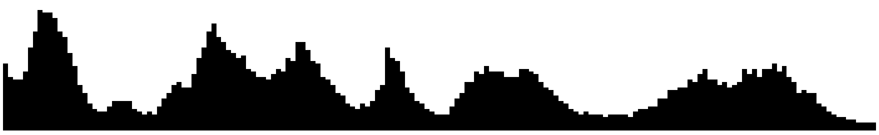
Variant A-1.1	*Chunks*	head movement + arm wave; arm wave; arm lift; concluding rotation
	Density	low; moderate; l..
	Extension	low; moderate; low
	Involvement	low; moderate; low
	Speed	low; moderate; low
	Force	weak
	Articulation	smooth
	Emphasis	
	QoM	
Variant A-1.2	*Chunks*	inward - outward (opening); downward rotation; pulling up/stretching upwards
	Density	low; moderate; low; moderate; low
	Extension	low; moderate; low; moderate
	Involvement	low; moderate; low; moderate; low
	Speed	low; mode..; low; moder..; low
	Force	weak
	Articulation	smooth
	Emphasis	
	QoM	
Variant A-1.3	*Chunks*	fall; sneak; push down; slow lift; flap; gather
	Density	low; moderate; low; moderate; low
	Extension	moderate; low; moderate; low
	Involvement	igh; moderate; low
	Speed	moderate; low; moderate; low
	Force	weak
	Articulation	smooth
	Emphasis	
	QoM	

Table 22. Table showing an overview of the analyses of the *Lento* variants performed by dancer A.

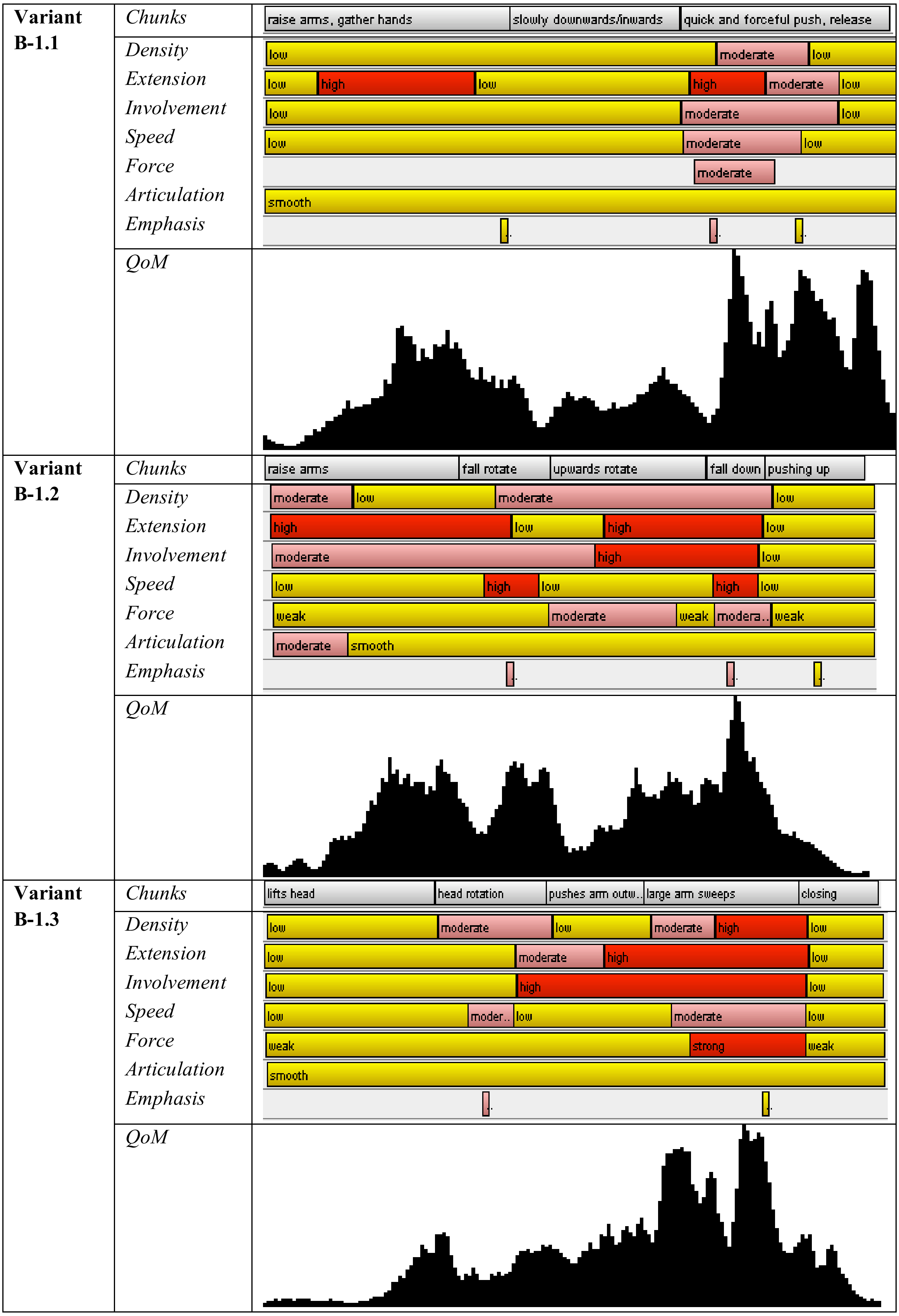


Table 23. Table showing overview of the analyses of the *Lento variants* performed by dancer B.

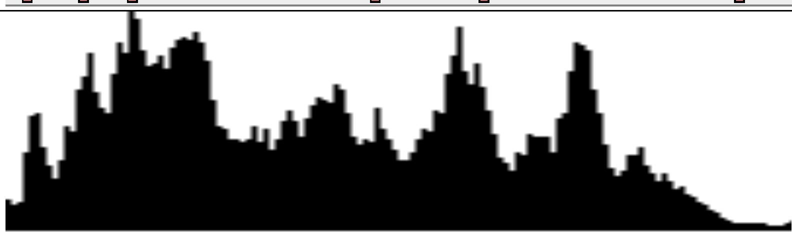
Variant A-2.1	*Chunks*	stret.. / flick / sweep / sweep / bow / curly rotations, arm sweeps
	Density	high / moderate / low
	Extension	moderate / low
	Involvement	high / low
	Speed	high / moderate / low
	Force	strong / weak
	Articulation	moderate / smooth
	Emphasis	
	QoM	
Variant A-2.2	*Chunks*	fan / heals / flex / circles / circles - bow / tugs / gliding
	Density	high / low
	Extension	high / moderate / low / moderate
	Involvement	high / low
	Speed	high / low
	Force	moderate / strong / weak
	Articulation	moderate / disctinct / smooth
	Emphasis	
	QoM	
Variant A-2.3	*Chunks*	stre.. / stret.. / st.. / sweeps / circles-flex / circles-st.. / weighted circles / slow arm raise
	Density	moderate / high / low
	Extension	moderate / high / moderate / low
	Involvement	moderate / high / modera.. / low
	Speed	high / low
	Force	weak / moderate / weak
	Articulation	disctinct / smooth
	Emphasis	
	QoM	

Table 24. Table showing overview of the analyses of the *Prestissimo variants* performed by dancer A.

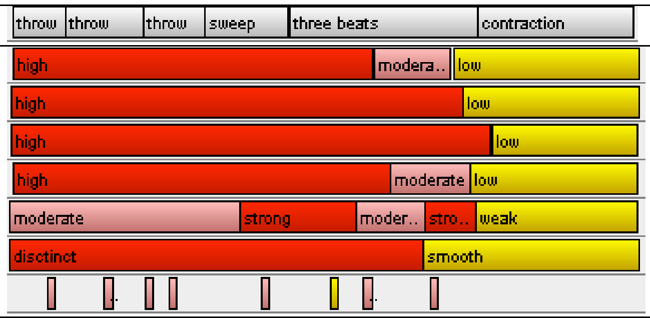
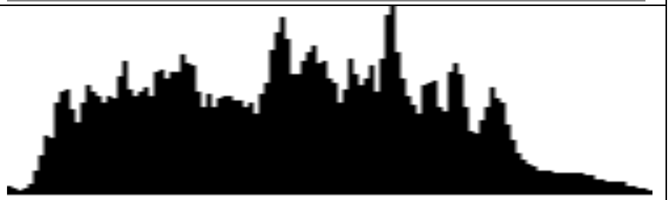
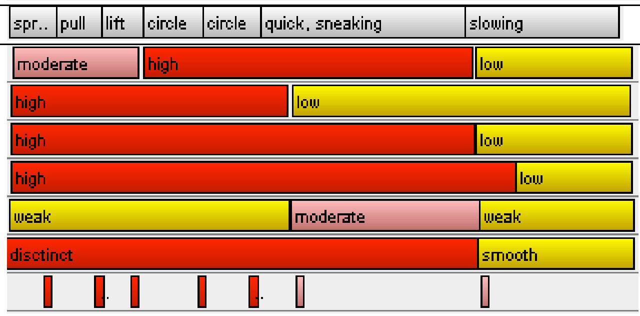
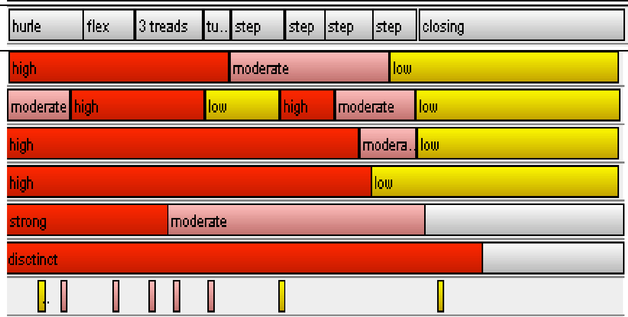
<table>
<tr>
<td rowspan="2">Variant B-2.1</td>
<td>Chunks
Density
Extension
Involvement
Speed
Force
Articulation
Emphasis</td>
<td>
</td>
</tr>
<tr>
<td>QoM</td>
<td></td>
</tr>
<tr>
<td rowspan="2">Variant B-2.2</td>
<td>Chunks
Density
Extension
Involvement
Speed
Force
Articulation
Emphasis</td>
<td>
</td>
</tr>
<tr>
<td>QoM</td>
<td></td>
</tr>
<tr>
<td rowspan="2">Variant B-2.3</td>
<td>Chunks
Density
Extension
Involvement
Speed
Force
Articulation
Emphasis</td>
<td>
</td>
</tr>
<tr>
<td>QoM</td>
<td></td>
</tr>
</table>

Table 25. Table showing an overview of the analyses of the *Prestissimo variants* performed by dancer B.

At first sight, the tables illustrate general differences between the *Lento* and *Prestissimo* variants; first with respect to chunking, and second in terms of articulation and the frequency of emphasis:

8.3.1. Chunking characteristics of *Lento* vs *Prestissimo* variants

There is a marked difference between the *Lento* and *Prestissimo* variants with respect to the number of chunks. The *Lento* variants are segmented in relatively few chunks (generally 2 - 4). Conversely, the *Prestissimo* variants are characterised by a high density of temporally short chunks at the beginning (4-5) and more stretched out chunks towards the end of the segments. (The annotations of the chunking features are based on the criteria explained above. Features contributing to chunking will be discussed in more detail for selected variants later in this chapter).

As was described in the analysis of the *Lento* the musical excerpt is phrased as two stretched-out waves where the finishing segment is more clearly articulated as a peak structure than the first wave. Considering the stretched-out chunking in most of the *Lento* variants, it seems that the dancers are attempting to respond to the phrase/chunk structure in the music.

Regarding the *Prestissimo*, the music is characterised by hasty, successive figures so that we to a lesser degree can identify clearly demarcated chunks. This means that it is not possible to see-hear chunk by chunk correspondence as we could in the *Lento* variants. Thus, I understand the high density of temporally short chunks in the movement variants more as a response to the general high level of activation, as though the dancers were trying to introduce 'newness' at a high rate in order to express the haste in the music. Accordingly, the more stretched-out chunks towards the end correspond to the decreasing level of activation in the music.

8.3.2. Differences with respect to overall articulation and emphases

In the tables one of the activation tracks indicates a moment of emphasis (accentuation), whereas another track describes overall articulation, i.e. whether the movements merge smoothly or are more distinctly articulated, for example characterised by abrupt changes in direction.

The annotations of the *Lento* variants show that the movements are generally soft and gliding, that there are few moments of emphasis, and that these moments tend in many cases to be only vaguely suggested (in yellow). The most distinct accentuations are found in the temporal segment where the music pushes forwards towards a goal-point, which suggests that the dancers are attempting to express the *peak structure* of the music.

In the *Prestissimo* variants on the other hand the movements are much more distinctly articulated. This is most markedly apparent in the first part of these gestural variants, and in this segment we can also typically see a succession of 'beats' or clear accentuations so that each chunk acquires a *peak structure*.

Towards the end of the *Prestissimo* variants the articulation becomes smoother and there are fewer and in some cases no moments of emphasis.

If we consider these aspects together, I would suggest that the density of chunks over time, the overall articulation feature as well as the distribution of emphases may be understood as responses to differences in the overall level of activation and the activation contour, i.e.:

- the *Lento* is characterised by a low level of activation and a slight increase/decrease in activation towards the end, and
- the *Prestissimo* is characterised by its high level of activation, and a change from a high level in the first part to a lower level.

These aspects are further indicated by the remaining activation features. In the *Lento* variants they are in most cases annotated in yellow (low level) at the beginning, and light red (moderate level) towards the end. Conversely, the activation features in the *Prestissimo* variants are shown with dark red (high level) at the beginning and with yellow (low level) towards the end. Finally, the QoM-analyses in most cases show the same tendencies when understood as an indication of changes in activation over time.

However, there are a few exceptions to these general tendencies. In the *Lento* variants A-1.3 and B-1.2 the Anvil annotations, and to some degree the QoM-analyses, suggest that there is a higher degree of activation in the first part. This means that the overall activation contour of these two variants is different from the other *Lento* variants and diverges accordingly from the activation contour of the music. Not unexpectedly, these two variants according to my judgement correspond to the music to a lesser degree than the other variants.

Similarly, the QoM-analyses of the *Prestissimo* variants A-2.3, B-2.1 and B-2.3 do not show a clear descending contour as in most of the other variants. But whereas a diverging QoM-analysis was an indication of a lower degree of match in the *Lento* variants, this is not the case here. The variant A-2.3 will be discussed in detail later.

In sum, tables 22 to 25 provide a basis for a preliminary discussion of correspondences primarily with respect to the general level of activation and the overall activation contour. From this rudimentary overview I shall proceed to a more detailed analysis. This will include:

- Judging how the degree of correspondence varies within one variant, as well as the features that contribute to these variations.
- A further description of activation and chunking

In addition I will discuss how aspects such as dynamics and the distribution of synch points affect correspondence. The term dynamics refers to aspects of movement described by metaphors such as squeezing, pressing, and floating, which were discussed in the previous chapter in connection with Laban's notion of basic effort actions.

For example, the *Lento* variants A-1.3 and B-1.2 just mentioned do not in my view diverge from the music solely in their activation contour. The movements in A-1.3 are generally very light and airy. In some segments they also appear to be slightly too quick and the final movement, characterised by a single slow flap of the right arm is very light and lacks the 'stretching', 'squeezing' features of the music. Similarly, the movements in B-1.2 seem to be too quick. In addition, they produce a sense of restlessness; they appear to be too unsettled and to have too much of a freely falling, releasing/unrestrained quality, and may at some moments give a too forceful and abrupt impression.

The following detailed descriptions of movement variants will demonstrate what has been commented on earlier, i.e. that the two dancers seem to use well-rehearsed movement patterns so that one sequence of movements builds on one or two 'ideas'. The discussion of correspondences should be seen in this light: one might in some cases say that the movement variant is more or less "inspired" or "guided" by this strategy and not solely based on changes in features in the music.

8.4. Detailed analysis of two *Lento* variants

I have selected first the *Lento* variants A-1.2 and B-1.1 for more detailed analysis because I judge them to be examples of good matching with the music.

8.4.1. Analysis of A-1.2

In variant A-1.2 the dancer starts off with her two hands gathered in front of the body. From this position her hands are pulled upwards to a position at chest level. The right hand is stretched forwards and the head and gaze are lifted upwards. She starts a rotation in her hips, the left arm stretches to the floor and the right arm stretches upwards. She finishes by pulling up with a movement initiated in her legs and closes the segment by stretching her right arm slowly upwards.

The movements of this variant are slowly and smoothly performed. It is thus similar to the general movement character of example A-1.1. Although evenly and lightly performed, the movements may be characterised as a little more heavily floating, expressing a more elastically stretched quality compared to A-1.1. The flow of the movements (in terms of dynamics) also seems to be more broken up, i.e. we have brief moments of light jerks/tugs, and we have moments at which the movements are more markedly and firmly emphasised/articulated. This supplies the movements with a more pulsating and vivid quality. This contrasts A-1.1, in which the stream drifts almost imperceptibly from one movement to the next.

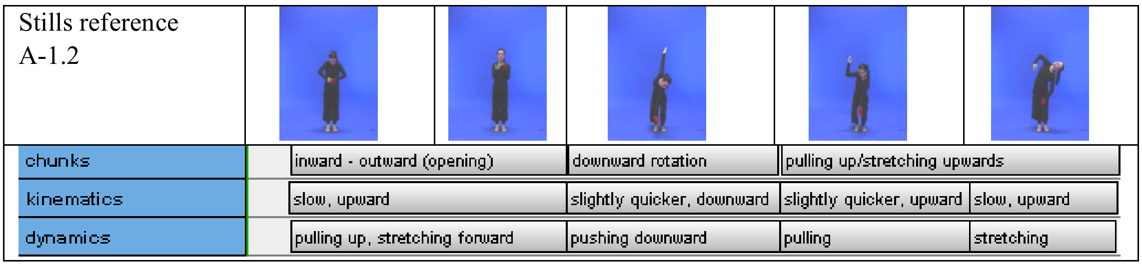


Figure 85. Selected stills from A-1.2 and an excerpt from the Anvil annotation showing descriptions of chunks, kinematics and dynamics.

Description of variations in correspondence

There is a high degree of correspondence throughout. The following sums up the features I believe contribute to the close link between music and movements in this variant:

- The movements are performed with a general smoothness and slowness, as described above.
- The metaphors of pulling, pushing and stretching used to describe the shadings of this performance suggest that a heavier and more sustained quality of force or tension characterises the variant. This quality captures the stretching/squeezing quality of the musical flux
- Transitions are very smooth, as is also the case in the music
- In the middle of the segment the right arm is pushed down to the floor and this corresponds to the onset of the first musical figure and the following long, stretched-out tone. The bottom position of the right hand is perceived as synchronised (moderately) with the first figure in the music.
- The final upwards stretch of the right arm is seen as corresponding to the stepwise musical figure and with the concluding release of tension. Within this segment an additional synch point emerges.

The correspondence is slightly weaker in the middle part of the segment. At a certain point it seems that the movement pauses, whereas the music moves forward so that the music and movement are perceived as being "out of synch".

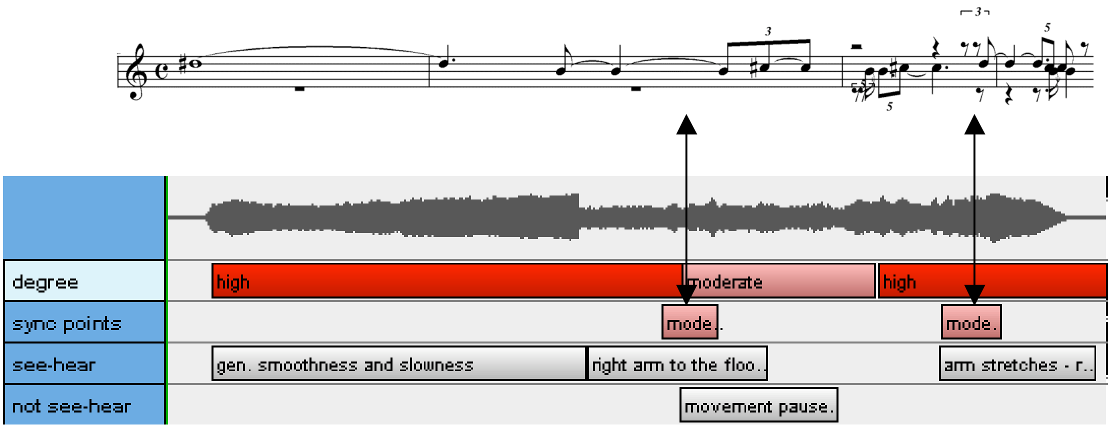


Figure 86. Excerpt from Anvil annotation showing judging of variations in correspondence with waveform and score excerpt as reference to the musical component. The arrows indicate the timing of synch points.

Activation contour correspondence

I would suggest that the overall activation contour may be characterised as a long, stretched, curve with a low initial activation level that is gradually increased, and then a more marked increase and decrease of activation in the final gestural segment.

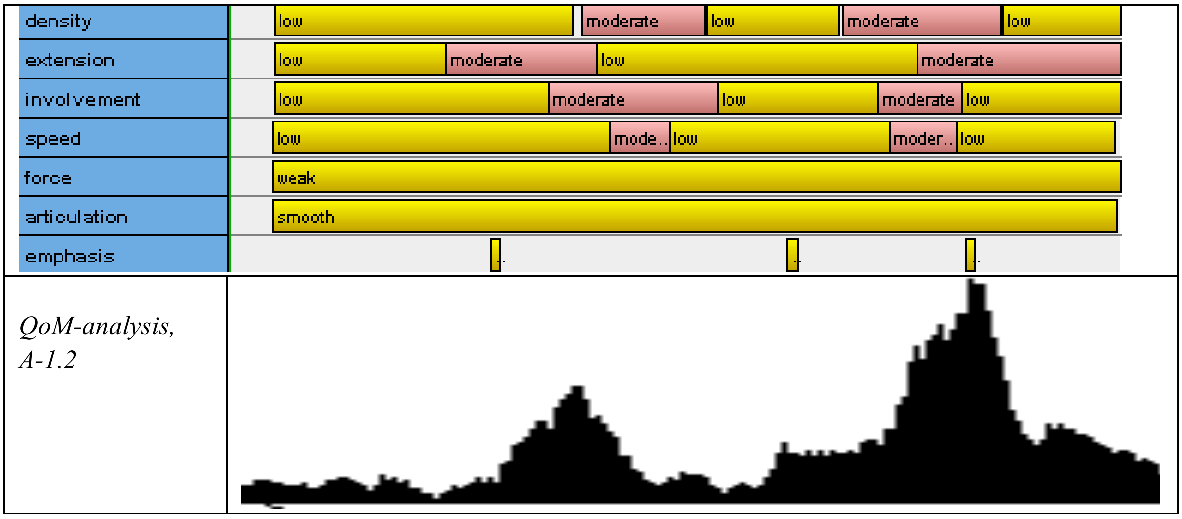


Figure 87. Excerpt from Anvil annotation showing judging of changes in activation features seen with QoM-analysis.

The Anvil annotation points in the direction that the variant is perceived with two increase/decrease pattern both resulting from changes in density, extension, involvement and tempo. The QoM-analysis suggests that the second of these two patterns is more marked, i.e. that the magnitude of change is more apparent.

Chunking/patterning correspondence

When observing the movements alone I have initially identified three main chunks: the first one lasts from the moment when the dancer has both her hands gathered before her to the moment at which her hand stretches forward and is characterised by a change from an introverted to an extroverted attitude (inwards/outwards). The two succeeding chunks are characterised by vertical directions; first, a push downwards towards the floor, and then a stretch upwards. There are also changes in the speed that may further contribute to patterning as summarised in figure 88:

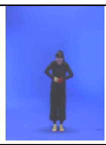
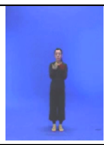
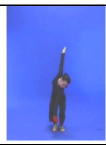
Stills A-1.2					
Variations and patterning of trajectory	inwards	outwards	down	up	up
Variations and patterning of speed contrasts	slow	quicker	slow	quicker	slow

Figure 88. Description of patterning of trajectory and speed contrasts.

When seen in combination with the music it appears that the inwards/outwards chunk and the rotation-downwards chunk in the movements merge so that together they correspond to the first wave in the music. Accordingly, the final figure in the music corresponds tightly to the third chunk of the movement sequence. This means that there is a close correspondence with regards to chunking. This does not only result from synchronicity in beginnings/endings and accentuation. It seems that there is an interplay between structural and dynamical aspects, for example between the timing of accentuation and the way accentuation is performed. I will come back to this point later.

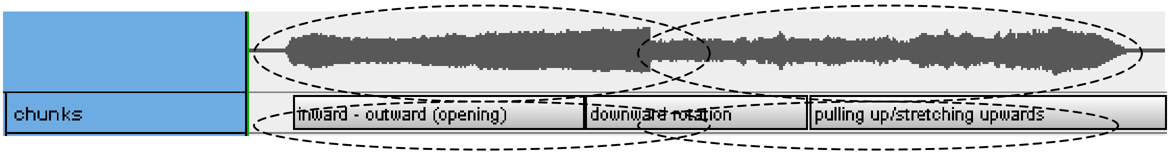


Figure 89. Excerpt of Anvil annotation with dashed circles suggesting a temporal correspondence between musical and gestural chunks.

8.4.2. Analysis of B-1.1

In variant B-1.1 the dancer starts by raising her arms slowly to a position above her head. After a brief rotation her hands meet in this position. And then, with her hands clenched together she pushes them slowly downwards. Finally, from a contracted posture with her hands held together, she forcefully pushes her arms out from her body before loosening up by flexing her knees lightly.

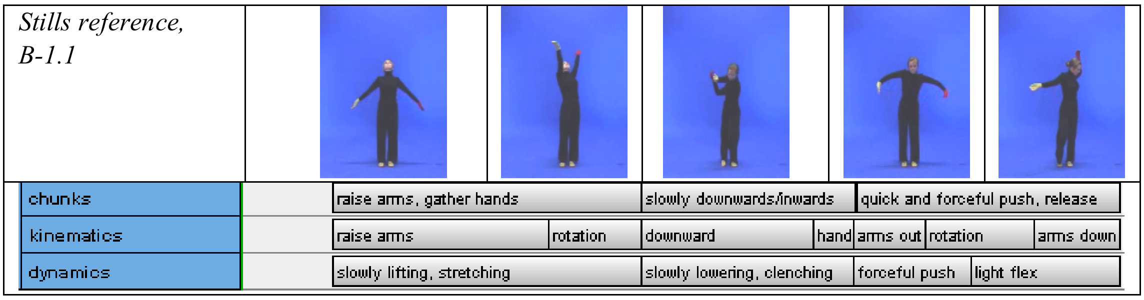


Figure 90. Selected stills from B-1.1 and excerpt from Anvil annotation showing descriptions of chunks, kinematics and dynamics.

The movements are characterised by being slow and gliding. The movements merge uninterruptedly and non-accentuatedly into one another. This contrasts markedly with the finishing segment in which there is a forceful push outwards as though she is pushing something heavy away from her body. This pushing movement appears to be abrupt compared to the gliding and gradual quality which came before. The movements are also characterised by a certain tension that runs through the stream of movements like a taut or stretched string. For example, this is seen in the downward movement with the hands clenched which produces a controlled, squeezing and detained impression. Another way of describing this is to refer to the general quality of dancer A’s style of movement which appears to be very light and easily performed as though she were floating on air. However, dancer B’s movements in this variant produce a slightly more controlled and forced impression which one might denote as lifted, detained, clenched, and pushed. This general character corresponds well to the tension in the music.

Description of variations in correspondence

There is tight correspondence throughout, one which becomes even tighter towards the end, as illustrated in the Anvil annotation in figure 91. The annotation also suggests moments of more moderate degrees of correspondence. These brief moments mainly result from a higher density in the movements than in the music.

The following summarises the initial assessment of aspects that contribute to correspondence:

- A sustained, calm and stretched nuance of movement that corresponds to the sustained timbre and smooth transitions in the music
- A weak synch-point occurs as the hands meet above the dancer's head: this corresponds to the initial tones of the first figures in music
- The 'clenching' movement downwards corresponds to the first stretched tone of the stepwise figure in the music
- Pushing both arms out to the side forms a clear synch-point with the accentuation in music (second tone in the stepwise figure)
- The release of tension which starts with a push towards the floor with the right foot corresponds to the release of tension in the stepwise figure in the music

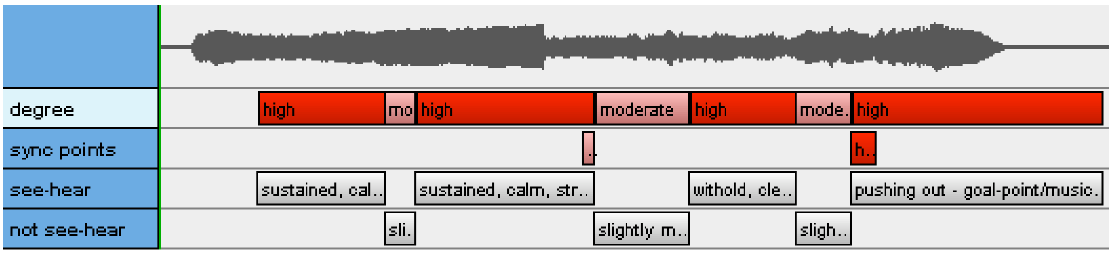


Figure 91. Excerpt from Anvil annotation showing judgement of variations in correspondence with a waveform as reference to the musical component.

Activation contour correspondence

The activation contour corresponds to the tempo changes, i.e. the activation is relatively low and non-changing at the beginning although a slight, momentary increase/decrease may be sensed in the rotation just before she starts the push downwards. There is then a marked increase in activation in the final segment, i.e. the forceful push outwards. The Anvil annotation shows that the qualities affecting activation are changed (increased) towards the end. Likewise, the QoM-analysis indicates a greater change in movement in this temporal segment. Thus, the activation contour of the movements is similar to the activation contour of the music.

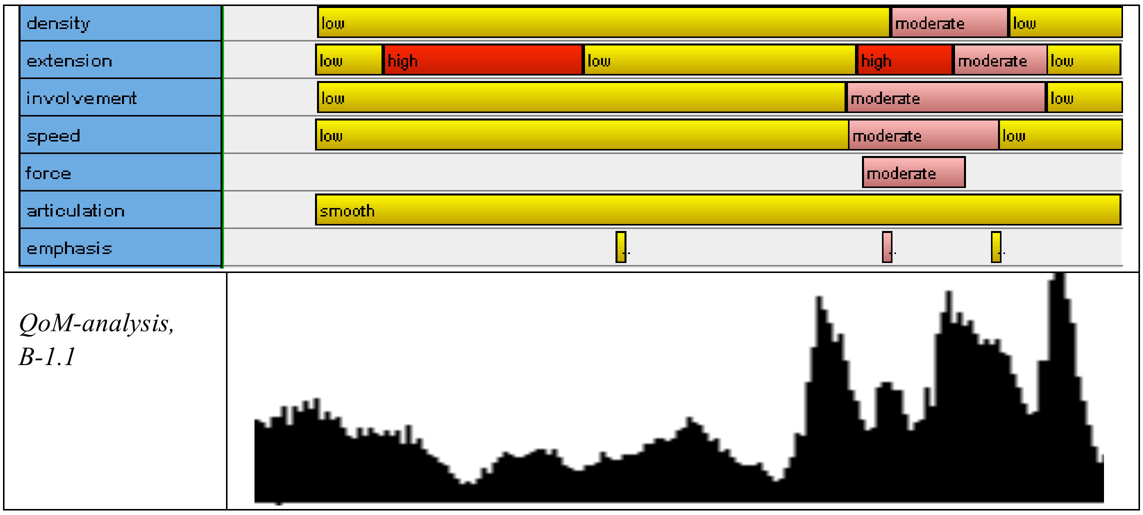


Figure 92. Excerpt from Anvil annotation showing judgement of changes in activation features seen together with QoM-analysis.

Chunking/patterning correspondence

As indicated by the Anvil annotation (figure 90), the movement sequence may be chunked into three parts. I would suggest that this chunking, as well as a clear sense of coherency, emerge on the basis of kinematical and dynamical patterns (also summarised in figure 91):

- Main directions: up (raise arms) – down (clenching) – out (pushing) – in (contracting).
- Changes in speed: slow – quick – slow.
- General shaping of tension: Gradual increase – marked increase - release of tension.

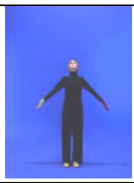
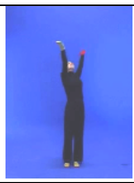
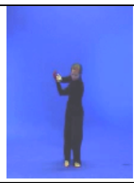
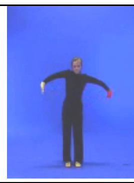
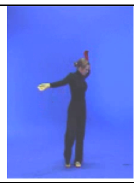
					
Main directions	up		down	out	in
Speed	slow			quick	slow
Tension	gradually increasing			marked increase	release

Figure 93. Patterns emerging from changes in directions, speed and tension.

It should be observed that the temporal distribution of chunks in this variant is quite similar to the one observed in variant A-1.2. Seen in combination, it would seem that the first movement chunk (raising the arms) plus the first part of the second movement chunk (slowly downwards) corresponds to the first wave in the music, whereas the second part of the second movement chunk and the third movement chunk (forceful push) corresponds to the second wave in the music. This means that in the combined condition the gestural flow is segmented slightly differently to afford a better match with the music. Interpreted this way, I would judge the correspondence between the music and movement with respect to chunking to be good, i.e. the beginnings and endings of musical and gestural segments are sufficiently timed, the synch-points are relatively clear and activation evolves similarly in the music and movement.

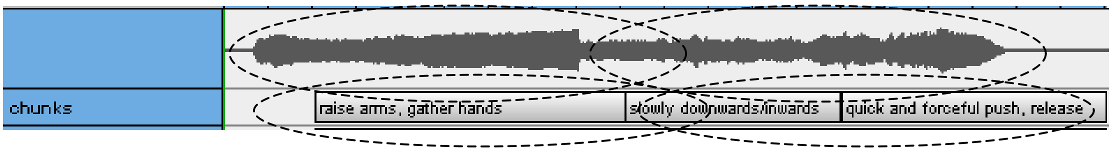


Figure 94. Chunking correspondence when observed in the combined condition.

8.4.3. Summing up correspondences in Lento variants and further discussion of correspondences with respect to dynamics and the timing of synch-points

The two variants A-1.2 and B-1.1 both demonstrate correspondence with respect to the activation contour as described in the previous section. They also show how the dancers relate to the chunking in the music, so that there is at least a coarse chunk-by-chunk correspondence between the music and movement.

Additionally, when I observe the *Lento* variants with the music there appear to be two aspects that affect my judgement of correspondence. These concern dynamics and the timing and quality of synch-points.

The first regards what may tentatively be described as the inner tension or a certain intimacy offered by the music. This feature has been suggested in my initial analysis of the *Lento* excerpt, i.e. the increasing brightness just before the two accentuations, and metaphorically described in the way the musical phrasing pushes forwards towards these accentuations first by pressing/squeezing and then by pulling/pushing. In a brief review using the waveform display as a timeline reference, these aspects may be temporally distributed as follows:

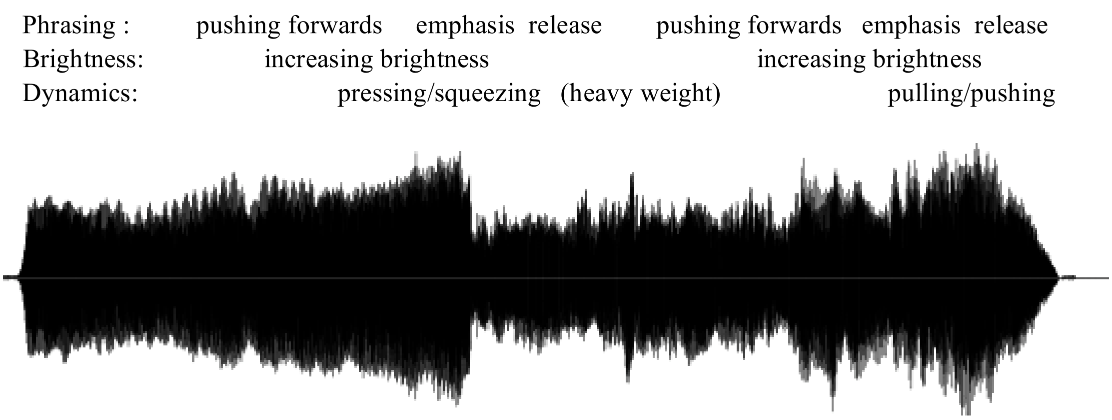


Figure 95. Features related to phrasing, brightness and dynamics in the *Lento* excerpt.

The dancers seem not to capture entirely this tension in the music; it is as though when the music is combined with the general smooth and airy quality of the movement variants, this aspect of the music becomes more apparent. This means that in the variants that I would judge to correspond more tightly with the music, the movements are performed so that they to a greater degree ‘absorb’ the sustained pushing/pressing/pulling quality that characterises the music.

Regarding the variants performed by dancer A, I think that they generally are performed with a very smoothly and lightly flowing airy quality. On the one hand, one might expect this smooth character to go well with the generally smooth and gentle articulation found in the music. But this smoothness seems to lack the pressing/squeezing/pushing quality that I think characterises the music. I judge A-1.2 to be closest to the music because this variant in my eyes to some degree captures this tension. The initial pull upwards with the two hands clenched together (image 1 in figure 96), as well as the pushing movement downwards with the right arm (image 2 in figure 96) that ends up in a quite firm *Weight* (Laban effort) both exemplify this. Another segment demonstrating this kind of convergence is found in the stretch upwards towards the end (image 3 in figure 96). In these segments the movement is characterised by a stretching, sustained quality, at least to the degree that they sufficiently converge with similar aspects of the musical flow.

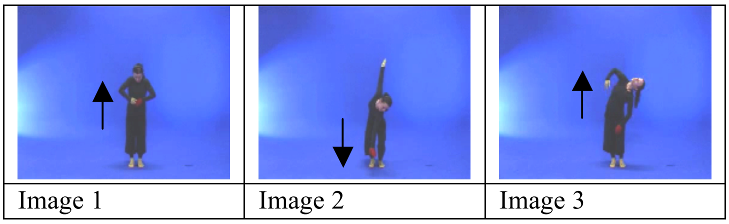


Figure 96. Three segments in variant A-1.2 that exemplify a pressing/pushing/pulling flowing quality.

Of the variants performed by dancer B, I would judge variant B-1.1 to match the music best. The detained slowness produces the impression of a pressing and lightly pushing quality that matches the music well. This is apparent in the way the dancer lifts her arms at the beginning (image 1 in figure 97), and in the segment where she lowers her clenched hands slowly (image 2 in figure 97), and finally in the final, more forceful pushing outwards with both arms (image 3 in figure 97). The dynamics of these movements also seem to match the dynamical shaping of the corresponding musical segments. This sustained, pressing quality is less consistent in the other two B-variants.

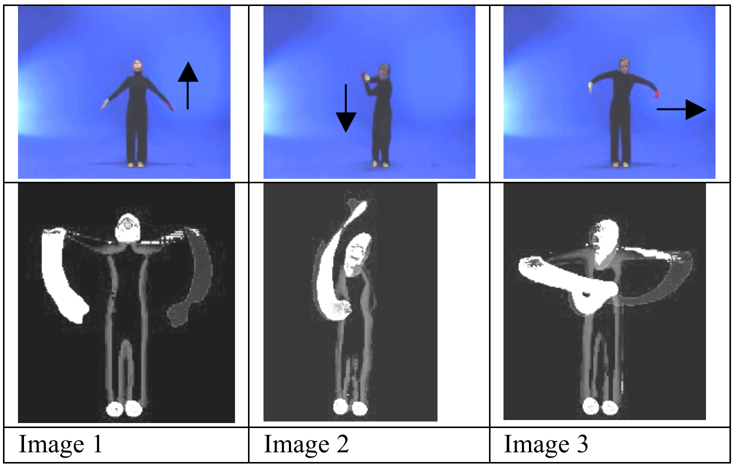
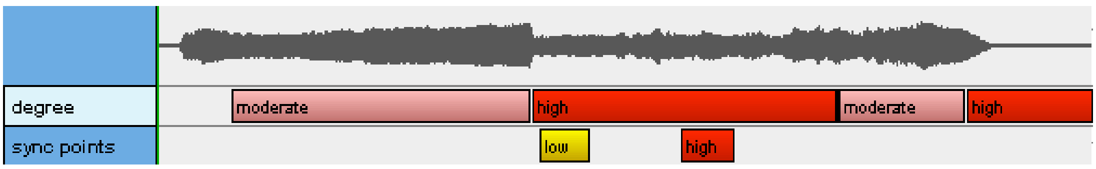


Figure 97. Three segments in the variant B-1.1 that exemplify using stills and motion history images a pressing/pushing/pulling flowing quality.

The other feature that defines the relationship between music and movement is the emergence of synch points, i.e. in some cases synch-points are quite strongly perceived whereas in other cases it is as though the music and movement were synchronised more loosely. Such differences are exemplified by the way the synch points in variants A-1.1 and A-1.2 are timed differently. In A-1.1 the two synch points appear slightly earlier than in A-1.2 (8s and 11s vs 10s and 16s), and this seems to make an important difference. The first synch-point in A-1.1 co-occurs with the onset of the first figure (after the initial sustained tone) in the music. In A-1.2 the synch point emerges just after the onset of this figure, as though the push downwards towards the floor coincided with the gently evolving musical figure. It is as though the quality of accentuation in the movement in the form of a sustained push rather than a momentary shove matched the quality of accentuation in the music. The differences in timing are illustrated in Anvil annotation of the two variants using the waveform display as a timeline reference.

A-1.1



A-1.2

Figure 98. Differences in the timing of synch-points in two *Lento* variants.

This illustrates that both timing and the quality of accentuation may affect the emergence of synch points. Furthermore, it seems that in the case of A-1.2 the temporal distribution of synch-points more effectively captures the phrase structure of the music.

Finally, I shall address an issue related to chunking correspondence demonstrated in variant A-1.1 which is characterised by successive arm lifts: the dancer starts by moving her head to the right and slightly downwards. Two wave-like right-arm movements follow this; the arm is twice lifted away and out from the body and moved back again to the side of the body. After the second wave, the right arm is lifted for a third time, but is now lifted higher. The arm is flexed when the elbow reaches the shoulder-height. From the top of the trajectory the arm falls in a rotating movement which also spreads to the rest of the body. The distribution of chunks is illustrated in Anvil annotation:

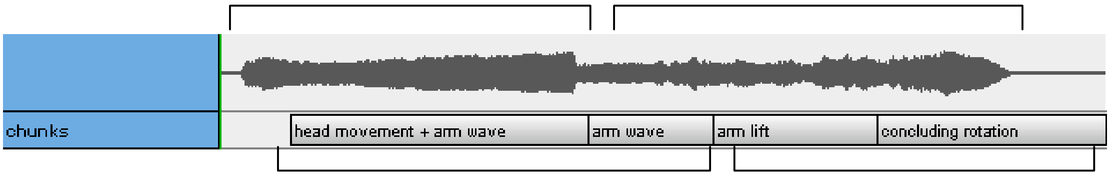


Figure 99. Distribution of chunks in movement variant A-1.1.

In my view, what one might term the kinematical theme suggested by the repeated lifts creates a coherency in the movements so that the two first lifts (i.e. the two waves of the arm) are seen as one coherent chunk. This is indicated in the Anvil annotation. Furthermore, considering the temporal distribution of chunks in the music, we can see that the musical and gestural chunking are slightly out of synch. One could perhaps have expected this to result in a sense of there being mismatch between the music and movement, but I do not think this is the case. On the contrary, it is as though the movements add an extra textural layer to the music, as though the differences in chunking appear as an interplay/collaboration between music and movement. This illustrates that the dancer through interpretation imposes a structure on the music in an attempt to anticipate a phrasing that matches the music more or less.

8.5. Detailed analyses of two *Prestissimo variants*

This tendency to impose a chunking pattern on the music is even more apparent in the *Prestissimo* variants. The short, leaping figures that characterise the music from the *Prestissimo* are not heard as distinct chunks; rather, the excerpt is heard as one coherent chunk. This chunk affords different changes in features (shapes), such as *quick – slow*, *distinct articulation – smooth articulation*, and *light – heavier*. On the contrary, the movement variants may be characterised as being chains of successive chunks that are clearly demarcated, relatively short at the beginning (1-2 seconds) and more stretched out towards the end (2-6 seconds). The temporally short chunks in the first part of the variants are also in most cases - most prominently observed in the B variants - characterised by emphasised moments so that a certain beat pattern emerges. However, these gestural beats are not synchronised with accentuations in the music. In light of this, it seems irrelevant to investigate the *Prestissimo* variants in terms of chunk-by-chunk correspondence. And accordingly, synch-points do not play an important role in the perception of correspondence in these variants.

Instead, I would propose that the dancers in different ways attempt to bring out the shapes suggested above that the music affords. I have chosen two variants for a more detailed description. They exemplify how the dancers interpret the overall shapes of the music. In addition, the two variants A-2.3 and B-2.2 are characterised by fine shadings that seem to affect the perception of correspondence.

8.5.1. Analysis of A-2.3

Similar to the other two Prestissimo variants performed by dancer A, the movement character (dynamics in Anvil annotation) of A-2.3 may be described as changing from light and quick through slower and heavier and then end up in a light and very slow arm-movement. However, the movements are more quickly and outwardly performed so that they appear brighter than the other variants.

The first part consists of three sudden stretches performed with a stiff body posture followed by quick, flapping arm sweeps. The initial stretches are distinctly and abruptly articulated whereas the sweeps have a gliding, non-articulated airy quality; she waves her arms/hands up and in front of the body as though she were flaunting a light ribbon in the air. After this follows three gestural shapes. They are linked together by their kinematics (trajectory), i.e. they are characterised by smoothly circling and slithering arm/hand-movements, as well as rotations of the body initiated in the hips. The sequence is finished off by raising the left arm lightly outwards.

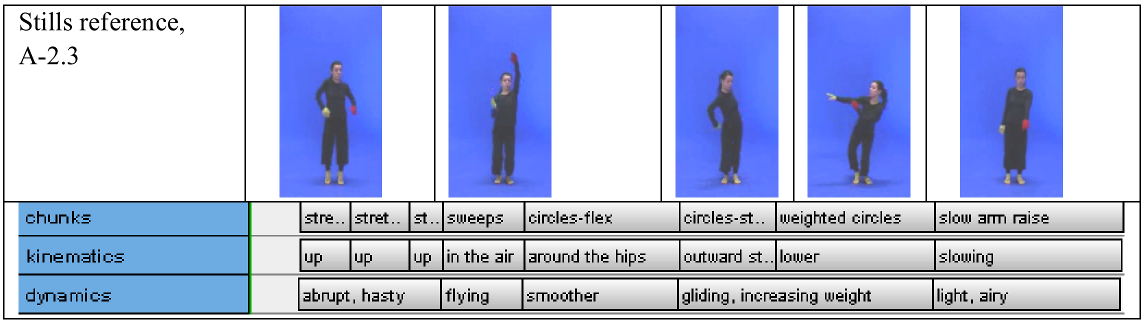


Figure 100. Selected stills from A-2.3 and excerpt from Anvil nnotation showing descriptions of chunks, kinematics and dynamics.

Judgement of variations in correspondence

I have annotated the degree of correspondence as high since I would judge the match to be generally higher than in the previous variants. The following aspects contribute to the judgement:

- There is an outwardness of attitude, in Laban's terms a more direct *Space* effort. This gives the movements a lighter and brighter character that seems to match the music well.
- In the middle part, the circling character of the movements seems to match the leaping and slithering musical figures.

- Furthermore, in this segment the movements become increasingly more heavily floating (firmer *Weight* effort) which corresponds to the heavier musical flow.
- In the final part the movements also slow down, similar to the music.

However, and most apparent at the beginning, there is a difference between the music and movements since the onset density is lower in the movements.

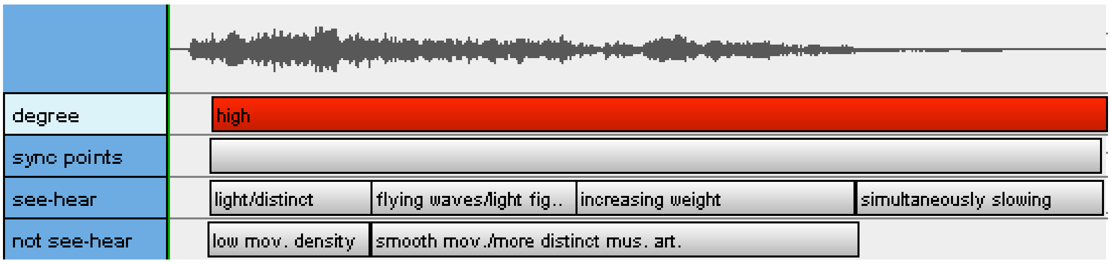


Figure 101. Excerpt from Anvil annotation showing judgement of the variations in correspondence with a waveform as a reference to the musical component.

Activation contour correspondence

Because of the lower density, extension and involvement (moderate) at the beginning, it seems that the level of activation in the movements does not quite capture the activation in the music from the start. However, the stretches are distinctly articulated (shown in the three sharp peaks in the QoM-analysis) so that the mode of articulation contributes to a sense of a high level of activation. The Anvil annotation suggests a relatively high level of activation in the middle part, whereas the QoM-analysis indicates a more moderate level. Despite these differences, I would judge the overall activation contour from a high to a lower level to match the music's activation contour.

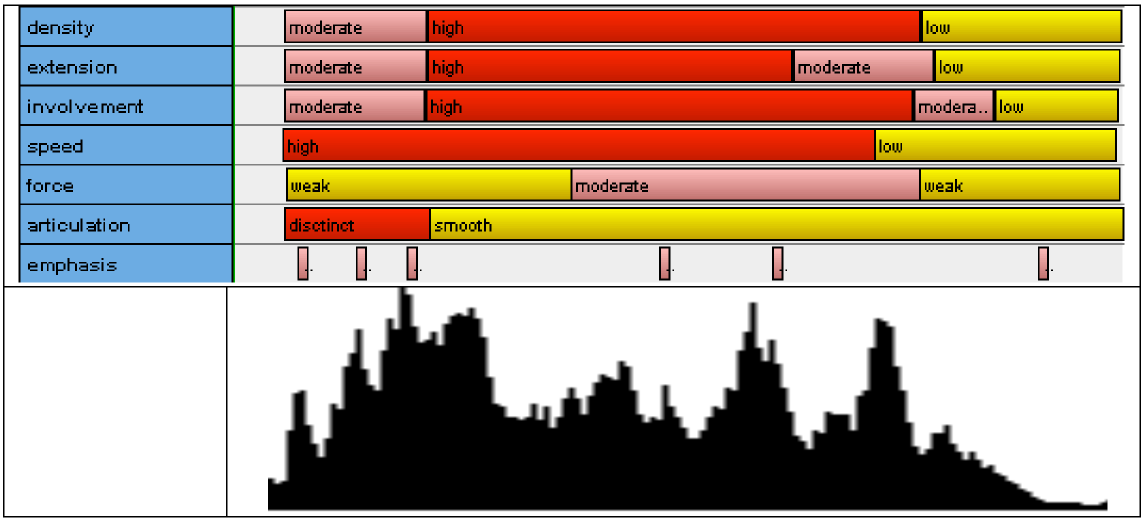


Figure 102. Excerpt from Anvil annotation showing judgement of changes in activation features seen together with QoM-analysis.

Correspondence with respect to dynamics

As mentioned in the initial description of the variant there seems to be a gradual change in the dynamics which may be characterised as a change from a comparatively light and quick *effort* to a more sustained and weighted *effort* (i.e. the term *effort* used in according to Laban Motion Analysis). This dynamical process seems to glue the movements together.

This change may be illustrated by the transition between the gestural shapes which may be characterised as patterns of circling hand-movements which end up in an act of stretching the arms horizontally out from the body (see figure 103):

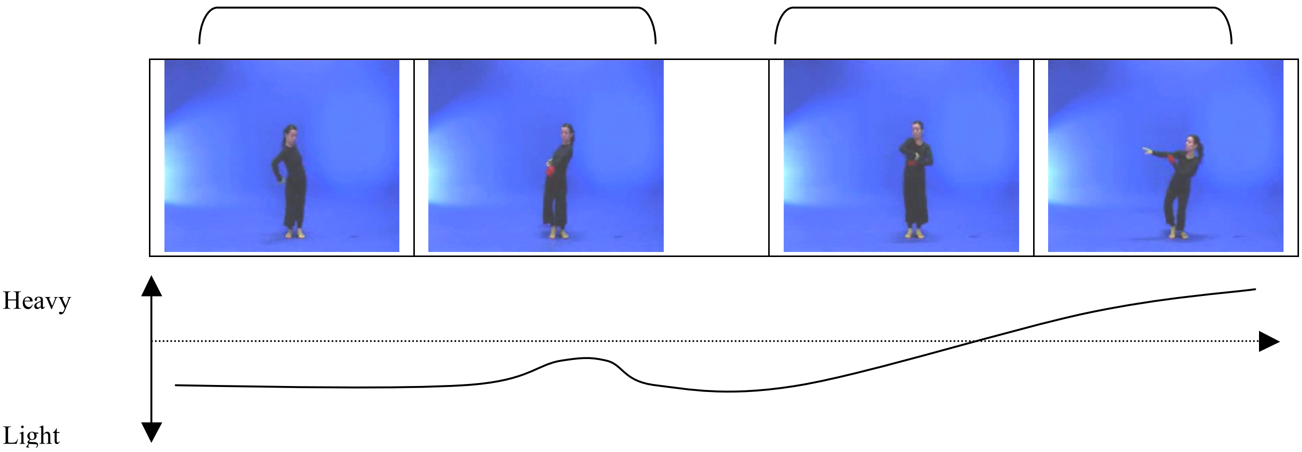


Figure 103. The figure suggests a gradual change from a light and quick effort to a more sustained and weighted effort in this movement sequence.

In the first of these shapes the stretch is distinctly and abruptly articulated with a quick bend of the knees. In the succeeding pattern the dancer is ‘sitting’ more deeply and the outward stretch has a heavier, more pressing quality, i.e. the applied weight/force is more drawn out in time. The curve illustrates the change from a relatively light to a heavier *Weight* (effort). The transition from this shape to the following sequence of gliding and slower rotations is very smooth. At the same time there is a change in articulation, from a light, distinct one to a more gliding, stretching one.

The dynamics are shaped by the following overall pattern which matches the *dynamical shaping* of the music, from light and hurried to more heavily floating:

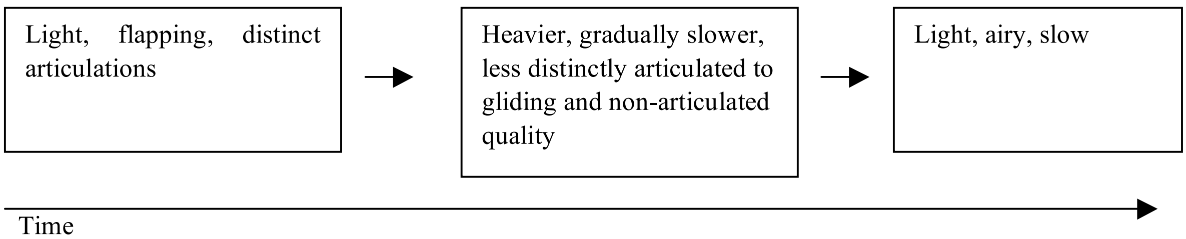


Figure 104. The change in dynamical shaping visualised schematically.

8.5.2. Analysis of B-2.2

As with dancer A, the variants performed by dancer B also demonstrate similarities. For example, all the B-variants are characterised by a general pattern of vertical lifting and lowering movements. They are also similar in the way the dancer touches the floor with her feet on tiptoe; she places her feet down in a controlled manner, i.e. she does not release her weight fully to the floor, as though she were placing her toes on a surface that is either too hot or too cold.

I have chosen variant B-2.2 for more detailed description since it demonstrates a nuance in the style of movement that clearly distinguishes it from B-2.1. The movements in B-2.2 appear to be performed more restrictedly and sharply. She keeps her body tighter. It is as though the space around her had been diminished so she were moving inside something narrow; she makes her body slender to squeeze herself into the narrow space. This slenderness and squeezing quality characterise the final segment of the variant.

As may be seen in figure 105 the first part is characterised by temporally short chunks which may be described as a sequence: *spreading the arms upwards - pulling upwards – pulling upwards – circling – circling.* The second part consists of two more stretched-out chunks in which the dancer pushes her arms/hands downwards close to her body, described in the Anvil annotation as sneaking.

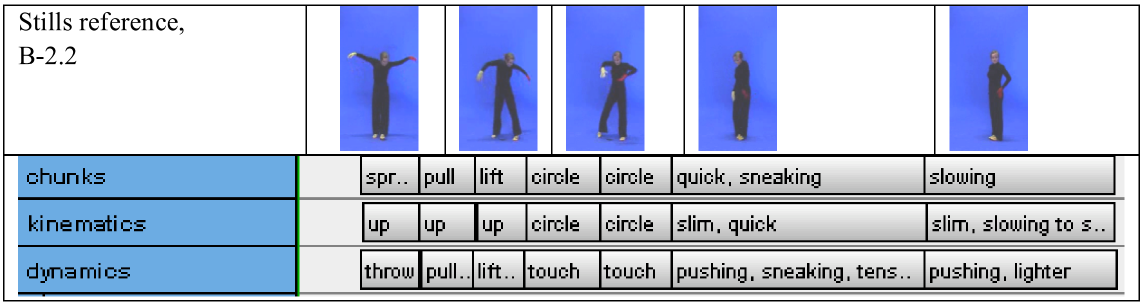


Figure 105. Selected stills from B-2.2 and excerpt from Anvil annotation showing descriptions of chunks, kinematics and dynamics.

Judgement of variations in correspondence

The degree of correspondence is judged to be moderate throughout. Generally, the onset density is higher in the music than in the movements. However, the movements are performed hastily and jerkily that matches the music well. Also, the sneaking quality towards the end appears to be similar to the changing character of the music in this segment.

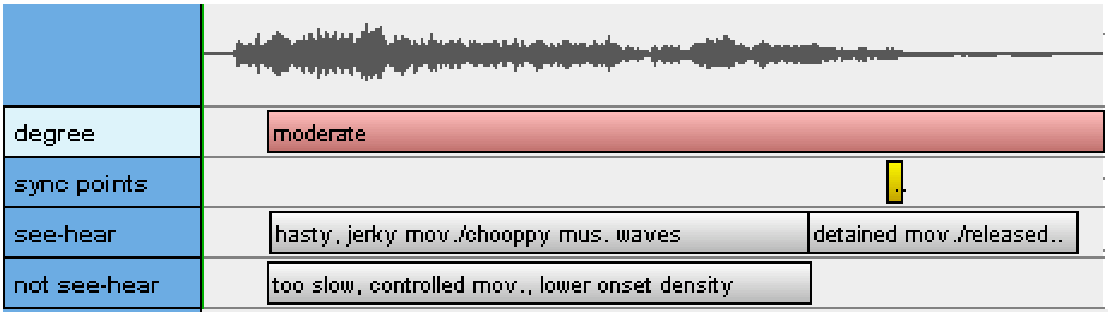


Figure 106. Excerpt from Anvil annotation showing judgement of the variations in correspondence with a waveform as reference to the musical component. The arrows indicate the timing of synch-points.

Activation contour

The Anvil annotation and QoM-analysis both illustrate that the activation changes from a high to a lower level which corresponds well with the activation contour in the music. The higher level of activation at the beginning results from a higher density of emphasized moments, a distinct mode of articulation, as well as contribution from the activation features' extension, involvement and speed.

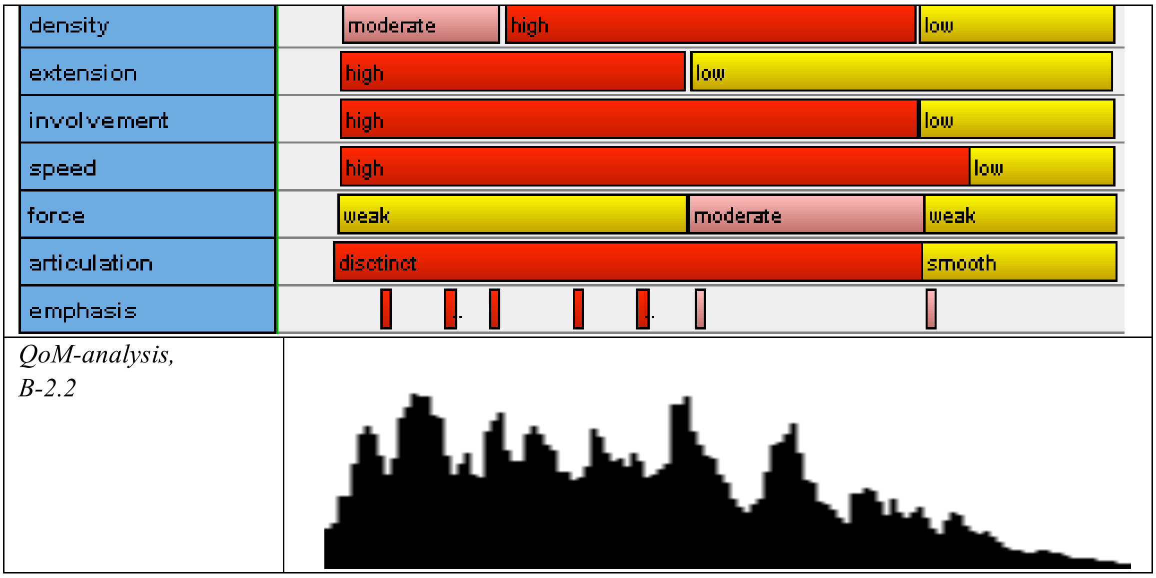


Figure 107. The anvil annotation suggests the way changes in the activation features extension, involvement and speed contributes to the decrease in overall activation.

Changes in dynamics

In line with the overall activation contour, there is a change in dynamics from quick, jerky, upward movements in the first part to intensely slithering/snaking and narrow, detained movements in the concluding part.

The movement character emerging from the movements upwards in the first part has a jerky, tugging character. It is as though she were trying to pull her body up and away from the floor which is seen in the way the dancer treads lightly. This character is apparent in the first gestural shape where the foot suddenly tugs away followed by a quick stretching movement upwards in which the arms shoot up into the air. The two succeeding shapes are more stretched, as though something is being pulled upwards sustainedly. After this follows two phrases that both consist of a circular plus a movement upwards. The circles are quickly and smoothly performed whereas the movements upwards are quick and airy as well as lightly and quickly articulated.

The first sequence of movements upwards is a contrast to the concluding sequence in which the dancer pushes her arms and hands down along the body, i.e. having previously been spread out from her body, the movements have now acquired a narrow and detained, crawling character. Combining the metaphors of ‘pushing’ and ‘sneaking’ describes the movement character quite well - although relatively short and quick, the movements have a tense and sustained quality. Within this last sequence there is also an overall change from a quick and tense to a slower and lighter character.

To sum up, the change in dynamics may be summarised as follows:

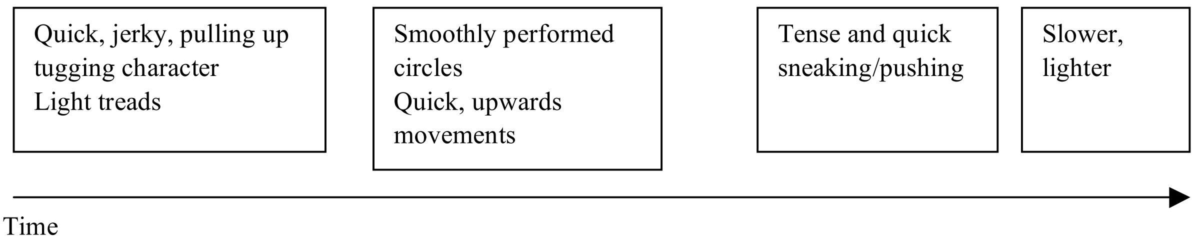


Figure 108. Change of dynamics in movement variant B-2.2.

8.5.3. Final comments on correspondences in the *Prestissimo* variants

To summarise, I think there are a few aspects that the *Prestissimo* variants have in common and which may be understood as the dancers’ attempts to capture changes in features in the music:

- First, and as mentioned earlier, all variants correspond well to the music with respect to its overall activation contour. This is apparent also in the more detailed analysed variants of A-2.3 and B-2.2.
- Second, I would suggest that the typical chunking pattern observed in the *Prestissimo* variants, from temporally short chunks at the beginning and more prolonged chunks towards the end, contributes to the sense of activation change.
- Third, although expressed in different ways, I think there is an overall change in the dynamics in all variants. These changes are described in table 26:

Variant	*First part*	*Middle part*	*Finishing part*
A-2.1	Quick, sustained, forceful, stretched-out, successive beats	Rotations, light, smooth, slower, more restricted	One long, slow, stretched movement
A-2.2	Light, stopping, quick, articulated with a light touch, abrupt changes in directions	More stretched-out, more forceful and more heavily articulated	Smooth, light glide
A-2.3	Light, flapping, distinctly articulated	Heavier, gradually slower, less distinctly articulated to a gliding non-articulated quality	Light, airy slow
B-2.1	Quick, bouncing throws. Alternation of heavy/long and light/short	Slowing, releasing character, decreasing activation	Contraction. Slow, inward, tensely clenching
B-2.2	Quick, jerky, pulling up. Tugging, light tread	Smoothly performed circles, quick upwards movements. Tensely and quickly sneaking.	Slower and lighter
B-2.3	Forceful to light, busy; bouncing	Weightless, floating, hurried	Slow, tender, fragile

Table 26. Table showing overview of descriptions of the way the *Prestissimo* changes from the beginning to the end.

These descriptions of change in features point in the direction that there are subtle nuances between the variants. In my view, these differences in the movements' character make a difference in the way correspondence with the music is perceived, i.e. in the way the movements capture the vividness of the music. I chose A-2.3 for a detailed description because I think the movements stay closer to the music than in any of the other variants. The movements are generally quicker; the turns are more sudden and abrupt which results in a lighter, brighter expression. The lightness of the movements can further be characterised with metaphors such as flittering, airy and fleeting, features that are found in the way she swings back and forth on her heels at the beginning as well as the quick wavy movements of the arms.

Conversely, the other two A-variants appear to be too smoothly articulated when seen in combination with the music. For example variant A-2.1 appears to lack the sharp articulation and sudden turns that are found in the music. The movements are evenly stretched. However, there is a forcefulness that seems to go well with the music, as though the strong, wriggling rotations make a heavy energy in the music become more apparent. In this variant the movements are taken more to the extreme in terms of extension, and this feature contrasts with variant A-2.2 which seems more detained. Thus, the movements in A-2.2 appear too soft and smooth as well as too controlled/bound. It is as though the movements lacked the nervous jerkiness that characterises the music. In this way the movements produce a "paler" impression.

There is one further aspect of variant A-2.3 that distinguishes it from the others and which I think makes a true difference. The dancer relates to the space around her more directly and outwardly, e.g. in the opening movement she stretches her body upwards with an open attitude towards space. True enough, the movement may not be characterised as a directly focused effort element according to Laban's motion factor of *Space*. The attitude to space is outward-looking and open but flexible/indirect in the sense of Laban's *Space*, and this differs markedly from the other variants in which the movements are much more inwardly directed. One might speculate that the closer correspondence is perceived partly on the basis of this feature, i.e. that the outward flexibility is similar to the way the music is performed, or alternatively that the movements bring out this feature in the music. This suggests that the music may be interpreted as having this outward, flexible quality. I think the image that I suggested in the initial musical analysis of the excerpt reflects aspects of this space feature: i.e. the image of someone who tries again and again to catch feathers in the air that escape outwards with quick, short, sudden movements and shift recurrently and flexibly. Certainly, it is difficult to identify exactly what it is in the musical performance that leads to this kind of impression. In the introductory analysis I have suggested that the imagery is brought out by the interplay between performative and musical features, such as unsystematic articulations (accents vs legato), non-periodic timing and pitch space characteristics.

Regarding the B-variants, I think the variant B-2.1 exemplifies a movement character that emerges too slowly and heavily. I miss a touch of wildness; the movements appear too controlled. As was suggested in the introductory analysis, the music may be heard as one long outburst of energy, i.e. a full breath is poured freely into the sound at the beginning, the intensity fades out gradually as there is no more air left. There is a stiffness and controlledness in the movements that subtly contradict the comparatively free flow in the music. It may have something to do with the breathing of the dancer, i.e. that watching her and empathetically imagining the movements gives the impression of something that is controlled and withheld. On the other hand there is a great, sweeping blow with the right arm that coincides with one of the flute figures, thus vaguely suggesting a synch point. The initial tiptoeing also has a light, jumpy quality that resembles the choppy, nervous turns in the music.

A slow and controlled quality also characterises B-2.2 but the movements in this variant have a more hurried, jerkier quality than in B-2.1. The movements are less extended in space and rounded-off. Nevertheless, the high density of onsets in music and the sudden and tugging/snatching quality contrast with the comparatively slow and controlled dance movements.

8.6. Summing up: flexibility vs non-arbitrariness

In this chapter I have discussed music-movement correspondences in six *Lento* and six *Prestissimo* variants. First, I commented on correspondences on the basis of chunking characteristics, activation features and QoM-analyses. Tables 22-25 illustrate clearly that the dancers responded differently to the *Lento* and the *Prestissimo* excerpts. The *Lento* variants, on the one hand, and the *Prestissimo* variants, on the other, are different with respect to certain aspects:

- General level of activation.
- Overall activation contour.
- Articulation.

Following this immediate impression, I selected two *Lento* and two *Prestissimo* variants for a more detailed analysis. In this part the variants were discussed both as they appear in the alone condition as well as how they appear when combined with music, and this included the following steps:

- A description of kinematics and dynamics on the basis of Anvil –annotation.
- A judgement of the way correspondences vary within one single variant (combined condition).
- A description of activation features contributing to overall activation contour.
- A discussion of correspondence with respect to chunking (*Lento* variants)
- Changes in dynamics (*Prestissimo* variants)

On the basis of this, I selected one theme for further discussion: I suggested that *dynamics*, i.e. the way the movements may be characterised by metaphors such as pulling, stretching, floating and gliding, and the way these shadings of performance change within one variant, are also important for the perception of correspondence:

- The *Lento* variants are typically performed slowly and smoothly but I would propose that variants that capture a certain tension (characterised as stretching/squeezing) afforded by the music are perceived to have a higher level of correspondence
- The *Prestissimo* variants are typically performed with an overall change (shaping) in dynamics from quick/hurried to slow/gliding. Among these variants there are also nuances so that the two variants that were selected for a detailed analysis were judged to capture the vividness of the music better than the other variants

The issue of *flexibility and non-arbitrariness* in music and movement relationships has been a recurrent theme in this thesis. In light of the preceding discussions, and with regard to the kind of empirical material that has been analysed, I would propose the following:

- *General level of activation* and *overall activation contour*

It seems that music and movement tend to correspond non-arbitrarily with respect to *the general level of activation and overall activation contour.*

- *Articulation* and *speed*

Considering that the *Lento* variants are performed smoothly and slowly, and that conversely the *Prestissimo* variants are performed quickly and distinctly, it seems reasonable to suggest that articulation and speed also contributeto the perception of correspondence as non-arbitrary aspects.

- *Kinematics*

Most apparent is that the movement variants differ in their kinematical patterns, i.e. when kinematics is understood as the trajectories and directions of a movement pattern. Whereas kinematics emerged as a robust feature for correspondence in the sound-tracing study, this was not the case in the dance-movement material, i.e. the dancers do not consistently respond to up/down pitch changes with up/down directions. This may have something to do with the musical excerpts as changes in pitch do not prominently characterise the *Lento* or *Prestissimo* excerpts. Another explanation might be that the trained dancers were approaching the task with well-rehearsed movement patterns largely characterised by kinematical aspects, so that these may override the vertical displacement suggested in the music. In sum, I would suggest that music and movement correspond flexibly in terms of kinematics (trajectories).

- *Dynamics*

On the one hand, since nuances of movement performance that have been referred to as dynamics also appear to play a prominent role in the perception of match, this aspect might be understood as a non-arbitrary aspect. Furthermore, the discussion of changes in dynamics that seems to characterise the *Prestissimo* variants points in the same direction, i.e. that the shaping of dynamics is non-arbitrary. On the other hand, these changes are performed differently in the variants. In sum, we might suggest that the shaping of dynamics over time, i.e. the way the movements change and leave a sense of a dynamical contour plays an important role for correspondence (non-arbitrary), but that music and movement correspond flexibly with respect to the way this overall change is expressed.

- *Chunking*

I have described segments in which music and movement converge so that the beginnings and endings of the chunks in the music and movement respectively are synchronised, and so that in the same segment music and movement have a similar structure in terms of *prefix – goal-point – suffix*. Not unexpectedly, this results in tight correspondence. However, I have also drawn attention to examples in which we do not have clear chunk-to-chunk correspondence, e.g. in which it seems that music and movement are slightly out of synch and from which we get the impression that music and movement relate to each other as independent, yet connected textural layers. In light of this, I would suggest that chunking correspondence may strongly affect a sense of correspondence but that it is not a necessary requirement.

8.6.1. Future directions for studies of free dance-movements

I have chosen to concentrate my analyses on a few examples. This approach resembling a case-study facilitates in-depth investigations of a few variants. The aim has been to describe music and movement correspondences from many angles. In this way I think my own thesis differs from many earlier studies, in which for example the observations/measurements have been limited to intensity profiles and the beginning/end detection of chunks (Krumhansl & Schenk, 1997; Vines et al., 2005). I think my own approach is worth pursuing, but it should perhaps be combined with a kind of approach that includes more participants and a larger number of musical examples. I would propose that future work should take two complementary directions:

- *A 'few cases' study*

which is similar to the approach that has been used in this chapter consisting of a limited number of musical examples and a limited number of participants (2-3) who are asked to improvise 2 – 3 gestural interpretations for each music example. The analysis should address the same aspects such as general activation level, the overall activation contour, chunking correspondence, the shaping of dynamics, etc.

Although combined with computer-based QoM-analysis, I have mainly based the discussion and annotations on my own subjective judgement. In future work it would be interesting, using the same observational categories that I have implemented in the Anvil annotation board, to invite a panel of researchers with different backgrounds to annotate manually the videos. This would be useful for two reasons: first, it would be interesting to see how and in terms of which aspects observers differed in their judgements of music-movement correspondences; second, it would provide a broader basis for assessing how correspondences are perceived in terms of *non-arbitrariness* and *flexibility.*

- *A 'many cases' approach*

It would also be useful to conduct a study that includes more music examples (10-20), as well as more participants (10 – 20). In this kind of study it would be too time-consuming to perform the analytical procedure that has been used in this chapter. The extent of qualitative judgements would have to be more limited, and for example only include annotations of chunking, articulation and moments of emphasis. The question is whether computer-based procedures could convincingly replace some of the qualitative categories.

Basically, I think the analysis should aim to provide an overview of the general level of activation, the overall activation contour and the chunking characteristics so that variants could be compared easily. In my view the set-up used in tables 22 – 25 visualises how gestural variants are similar and different as responses to musical examples. A kind of comparison including more movement variants would provide a broader basis for further discussions of the non-arbitrariness of music-movement relationships.

Regarding musical examples, I would suggest that the criteria used in the follow-up-study would be worth pursuing, and this would in my view be relevant to approaches using both a few cases and many cases. First, musical excerpts that are mainly characterised by changes in a limited number of features seem to be easier to analyse/describe so that a comparison with movement responses appears more straightforward. Second, if for example we have a number of musical examples that are characterised by changes in density of onsets it might be possible to assess whether the movement responses are also consistently similarly characterised by changes in the density of onsets.

There is one further possibility: to acquire even more control by using musical or music-like excerpts designed by sound-synthesis. This approach would facilitate the systematic variation of the musical component so that the same movement variant could be combined with similar but slightly varied musical excerpts. This way of collecting audio-visual material would be appropriate for use in a match/mismatch study, i.e. by having a number of participants judge the level of correspondence in a number of audio-visual compounds. A match/mismatch study of this kind could for example inform us about the way respondents judge correspondence when articulation in music is changed (e.g. becoming more distinct) while the articulation of movement remains unchanged (e.g. smooth throughout). Thus, it might be possible to assess whether any of the activation features appear to play a more prominent role in the perception of correspondences than others.

8.7. Methodological issues in studies of music-movement relations

Systematic research into a phenomenon such as music-movement correspondence raises questions about the audio-visual instances that are chosen to represent the phenomenon as well as how to 'measure' or analyse these audio-visual instances most reliably. Above I have suggested future directions for studies of free dance-movements. In the following I shall discuss in more detail how the material from the sound-tracings and the free dance-movements together raise two core methodological issues which will be important to consider in future work:

- First, the main research question addresses how certain features in music interact with certain features in movement. This presupposes that it is possible to find a certain level of *consistency* in the way people in general perceive music-movement correspondences; I have referred to this as the *non-arbitrariness* of music-movement correspondences.
- Second, I have chosen sound-tracings and free dance-movements to represent the broader general phenomenon of *music-movement correspondences.* The question is whether the relatively sparse collection of music-movement instances is reasonably generalisable to other instances of music and movement. This raises the question of *validity*.

8.7.1. Consistency of correspondence judgements

In chapter 6 I attempted to evaluate consistency in sound/drawing correspondences by employing an interrater design. I examined whether twenty raters would make similar judgements of music-movement correspondences. The data analysis resulted in low Kappa values which indicate a low level of interrater reliability. Within the framework of quantitative research this would lead to the conclusion that the study does not support the hypothesis that sound/drawing correspondences are perceived consistently between many subjects.

Another way of testing consistency would be to find out whether one sound-tracer performed a similar drawing to the same sound at different points in time, i.e. testing reliability through a testing/retesting design. I suspect that we would see many examples of sound-tracers making slightly different and perhaps in some cases completely different drawings if asked to go through the same procedure again, for example a few weeks later.

How could we interpret this? It seems obvious to conclude that sound/drawing correspondences appear more arbitrary than consistent. On the other hand, we may again point to sources of inconsistency related to features inherent to the phenomenon, as well as features related to how the phenomenon is perceived (see also the discussion of sources of agreement/disagreement in chapter 6):

- First, considering the multidimensional nature of both sound and movement, it is reasonable to believe that the same tracer might address one feature (e.g. pitch) in the first test and another (e.g. timbre) in the re-testing so that the drawing would turn out quite differently at two different points in time.
- Second, there is the question of *sameness* vs *similarity*. These are two different things. I would suggest that the same rater might draw slightly different drawings on two occasions which strictly speaking in terms of quantitative measurement are not the *same* but which are perceived as similar or sufficiently similar so that they might be judged as the same.

This last point has been broached earlier on in the thesis. To illustrate this, I shall present four different but similar drawings (see chapter 3):

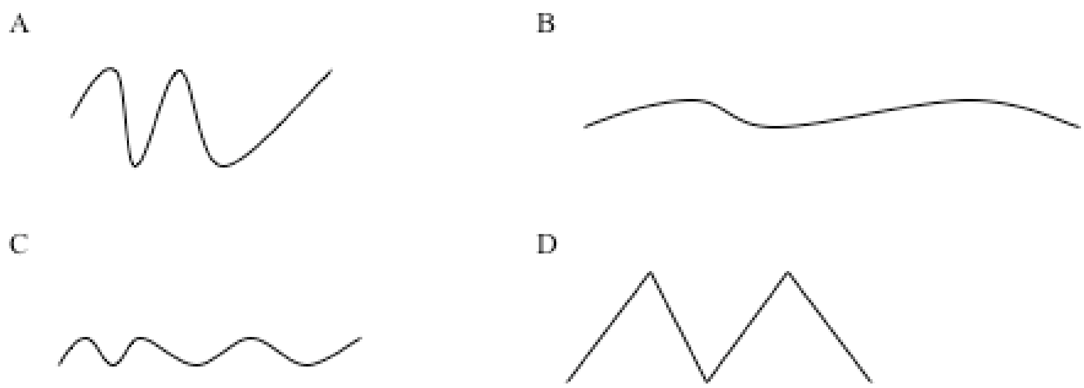


Figure 109. The shapes may be perceived as both similar and different depending on the perspective and context.

If asked, we would all agree that the shapes are not the *same*; but, if asked whether they are *similar* or *different*, I would suggest that most of us would say that they are both *similar* and *different*. This exemplifies our capability and tendency to assign different objects or events to the same category on an approximate basis and not in a fixed manner (Harnad, 1987), and that the process of perceiving similarity does not depend on features alone but also on context and the perspective we choose.

In chapter 3, the shapes above were used to illustrate how simultaneously evolving music and movement may be different in terms of specific features yet still sufficiently similar to be perceived as corresponding. Here, with regard to consistency in drawings as a response to sound, the theoretical perspective suggests that two different drawings meant to correspond to the same sound, either performed at two different points in time by the same sound-tracer or by two different sound-tracers might be perceived as similar although they are not exactly the same. If we look again at the display in figure 110 showing nine different drawings each meant to correspond to a trumpet sound, the question is: *how different can two drawings be and still be perceived as similar?*

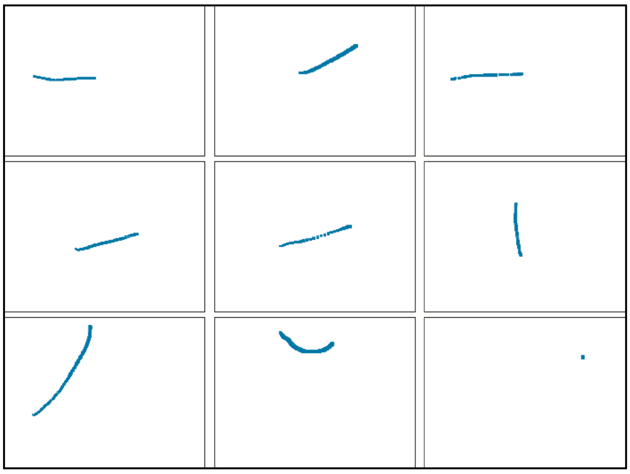


Figure 110. The drawings meant to correspond to a trumpet sound may be perceived as similar, except for the dot drawing, although they are not the same when quantitatively measured.

Apart from the dot in the bottom right corner, I would say that all the drawings are similar, and when the context is added, i.e. when given the information that they are all meant to correspond to the same sound, the judgement of similarity is strengthened. This means that the qualitative judgement of similarity, at least with regards to this specific example, supports the assumption that sound/drawing correspondences are perceived non-arbitrarily, whereas the quantitative approach, the inter-rater study with the calculation of Kappa values, does not.

The same applies to the different movement variants in the free dance-movements task. The variants are different but they also demonstrate similarities. The question again is: *how similar do they have to be to be considered sufficiently similar?* The question is apparent when we watch the variants in succession and on 'face value' judge them as similar, e.g. on the basis of slowness and a continuously applied force in (most of) the *Lento* variants and the hastiness (hurriedness?) of the *Prestissimo* variants. The point is further illustrated by comparing the QoM-analyses:

Figure 111. The two QoM-analyses of movement variants intended to correspond to the musical excerpt are regarded as similar although not exactly the same.

The two analyses are both similar and different, and since we know that they have been conducted on movement variants performed to the same musical excerpt, we probably impose a higher degree of similarity than we otherwise would.

In the analytical work I have moved from a quantitative approach (interrater consistency estimated by Cohen's Kappa) to qualitative analyses of individual sound-movement instances. For the present, it seems to me that a qualitative mode of 'measurement' provides the most reliable way of demonstrating consistency in music-movement correspondences. This suggests that quantitative approaches tend to miss the point, meaning that they do not capture the flexibility of perception.

In chapter 6, I discussed alternative ways of measuring consistency, e.g. by basing the rating procedure on an interval scale. This might or might not have produced a better result, but as long as we use sound material that is characterised by two or three co-evolving features I believe it is difficult to quantise consistency in music-sound correspondences. Another key issue is that we lack a kind of tool that is able to quantise similarity in a perceptually valid manner, i.e. essentially to judge among similar and dissimilar within a given context as our perceptual system does.

8.7.2. Validity of sound-tracings and free dance-movements

The other issue raised in this section regards the validity of the empirical material, i.e. does the choice of sound/movement instances provide a valid basis for making assumptions about music-movement correspondences in general? This is a question of possibilities and limitations. The sound-tracings and the free dance-movements have the following advantages:

- Excerpts of a short duration, facilitating a highly detailed analytical focus.
- Restricted in terms of dimensionality, i.e. in the number of co-evolving features (this is most clearly the case with the sound-tracings), hence, there is a focus with respect to the way they are characterised by changing features.

- In the sound-tracings the movement component is represented by 'frozen' drawings which enable a quick, at-one-glance judgement of the correspondence of nine drawings.
- The movement component in the sound-tracings is represented two-dimensionally and partly stripped of timing features, thus providing a simplification to enable a more focused analysis.
- In the free dance-movements material we have real music and real body movement though these are restricted in terms of the duration, style and feature characteristics, again, to enable a focused analysis
- The movement in the free dance-movements material is performed on the basis of an instruction that restricts the use of space (the participants were asked to remain on one spot), and the analysis is based on a fixed camera position. These recording-features facilitate a comparison between movement variants.
- The free dance-movements material is characterised by simultaneously evolving visual and auditory streams which on the one hand make analysis more difficult compared to the sound-tracings, but at the same time also make the audio-visual instance more real as a multimedia/multisensory phenomenon.

Certainly, the strengths of the material may at the same time be viewed as weaknesses. Regarding the sound-tracings, it is uncertain whether a drawing truly represents a body movement in the minds of the sound-tracers and the raters so that we should be careful about generalising from *sound/drawing* correspondence to *sound/movement* correspondence. On the other hand, it has been proposed in this thesis that music and movement are perceptually explored by *imposing* or *extracting* shapes so that such imagery provides a basis for solidifying time-dependent phenomena as chunks. If we appreciate this understanding, the sound-tracings as reflections of perceived shapes may be a valid representation of the way we perceive movement. However, there are more serious limitations to the sound-tracings regarding timing/dynamics and interaction effects:

- The 'frozen' movement format excludes timing features which are certainly a key characteristic of movement. This means that the drawing format does not capture the dynamics of movement. This is in my view a serious limitation given that the dynamics seem to contribute prominently to correspondence. This problem was also demonstrated in the sound-tracing study where it appeared that it was more difficult to articulate dynamical features than kinematical ones in the restricted two-dimensional format.
- Additionally, the 'frozen' format of the sound-tracings omits possible interaction effects which are assumed to be a core feature of multimedia/multisensory phenomena.

The sounds used in the sound-tracing study are just single sounds and certainly not music evolving over time, and the musical excerpts in the free dance-movements material are restricted in terms of musical style so that we should be careful before generalising from correspondences between *sound/drawing* and *Ligeti/movement* respectively to *music-movement* correspondences. Further limitations in the free dance-movements concern the training level of the participants as well as the restricted focus on *non-symbolic* aspects:

- The detailed analyses of the free dance-movements have been made based on videos of trained dancers, who, it seemed, were applying a certain strategy to the task. The movement variants may therefore exemplify a certain kind of correspondence performed by a small group of trained people with a certain kind of expertise. Although we did include videos of novices, these have not been systematically analysed in this thesis. This suggests that the analyses conducted in chapters 7 and 8 provide a restricted picture of *music-movement correspondences*.
- Lastly, the analyses have concentrated on the way correspondences emerge as a result of the *non-symbolic* aspects of meaning. This is also a considerable simplification since it is clear that *music-movement* correspondences within the context of a multimedia performance are characterised by the way the *non-symbolic* and *symbolic* interact and interfere with each other to produce meaning.

Seemingly, we have a trade-off between possibilities and limitations. In my view, it seems difficult, if not impossible, to 'measure' or qualitatively analyse the phenomenon on a sufficiently detailed level without squeezing it into a reduced format. I would strongly suggest that to capture features that are relevant for *music-movement correspondences* as a phenomenon performing the analyses on a highly detailed level is required. These details concern all kinds of *shapes* that may characterise music and movement respectively, and moreover, analysis should attempt to describe how co-evolving features interact. Hence, I would propose that we have to find a balance between audio-visual instances that is rich enough to represent at least to a certain degree the multidimensional, and preferably, the multisensory nature of music-movement relations, but which is also restricted enough so that the amount of detail does not become impossible to handle systematically.

To overcome the limitations in terms of generalisability of the restricted approach proposed here, I believe one could perform a series of investigations during which each source of information about music-movement relations was concentrated on a limited set of perspectives/features, i.e. combining and alternating between different features, such as:

- duration (short vs long excerpts)
- level of training (novice vs expert)
- sound category (everyday vs musical)
- musical style (periodical vs non-periodical, music characterised by a rhythmical meter vs music with no rhythmical metre)
- setting (laboratory-design vs live performance of music/dance)
- dance style (folk vs contemporary dance)

Chapter 9. Studying music-movement correspondences

9.1. Introduction

This thesis has attempted to contribute to an understanding of music-movement relations as follows:

- It describes and demonstrates how specific features in music and movement interact so that the two streams are perceived as corresponding.
- It discusses and exemplifies procedures and descriptive terms that may be applied to an analysis of music and movement.
- It elaborates on the notion of *non-symbolic* aspects of meaning, i.e. the way music and movement become meaningful by virtue of our response.
- It provides a theoretical and analytical framework for understanding the music-movement relationship as intimately linked as a perceptual reality.

In this final chapter I shall summarise the theoretical perspectives that underlie the thesis as well as provide a brief summary of the way the thesis answers the main research questions.

9.2. Theoretical threads

A few intertwining theoretical threads have run through the thesis with the purpose of providing a general understanding of the intimate link between music and human-body movement.

9.2.1. Multisensory research, motor theory and ecological theory

The theoretical framework builds on *motor theory* and theory of *multisensory perception*. Certainly, these two approaches are interrelated since our ability to imagine (or covertly simulate) the movements (or movement features) that might be involved in music performance must be rooted in some kind of multisensory resource of the perceptual system. There are basically two assumptions about music-movement correspondences that emerge from these theoretical perspectives:

- The responses to sound, observed in the studies of sound-tracings and the free dance movement may be viewed as gestural explorations of the sound as it evolves over time, where *exploration* means that the performer of movement 'picks up on' features in the sound that may be more or less matched by or expressed in corresponding movement features.
- The fundamentally multisensory features of our perceptual system imply that we have a vast knowledge of the way sound and movement work together so that they constitute multi-channelled events; in other words we are sensitive to features in movement that correspond or converge with certain features in sound. However, it also seems that the perceptual system is capable of putting things together, e.g. visual and auditory "somethings" which do not necessarily correspond well to

each other in certain features but which are perceived as corresponding because they simply appear at the same time at the same location. This suggests that correspondence is perceived partly on the basis of similar features in music and movement but that correspondence also should be understood as an *emergent feature* of multisensory perception.

In theory, one might think of a situation in which the two streams when observed separately are not experienced as similar, but converge when they occur simultaneously. In line with this, it has been shown by way of experiments that incongruent music and dance were judged to match to a greater degree when music and dance-movements were presented in combination than when they occurred separately (Mitchell & Gallaher, 2001). Certainly, such interaction effects (by Mitchell & Gallaher referred to as 'capture effects') should be considered when analysing music-movement relations. However, I believe that the analyses of the empirical material of this thesis demonstrate that similarity plays a central role in the perception of correspondences and that it is highly relevant to study such correspondences in terms of specific co-evolving features in music and movement.

The broader framework of the *ecological theory of perception* contributes with the following perspectives:

- The perceiver is in a reciprocal relationship with the environment and this is captured in Gibson's key concept of *affordance*. I have used the terms *extracting* and *imposing* to apply this understanding of reciprocity, i.e. when the dancer explores the musical sound in terms of movement she extracts features *from* and imposes features *on* the sound.
- The statement "to perceive is to understand" implies that we know something about the environment, or more precisely, about our relation to objects and events, and this knowledge constrains perception.

9.2.2. Analytical concepts

The notion of perceiving and understanding leads to the theoretical perspectives that may be referred to as *symbolic vs non-symbolic aspects of meaning*. Initially, I stated that the discussion of correspondences would concentrate on *non-symbolic* aspects of meaning. Hence, one kind of meaning/understanding that affects perception is *non-symbolic*, which is understood as an aspect of meaning that acquires its content through our immediate and sensitive response rather than through convention as is the case with *symbolic* aspects of meaning. To put it simply, the time-based phenomenon becomes meaningful *non-symbolically* in the way we perceive and respond to changes over time, i.e. the event is characterised by changes in features. In this regard I have attempted to identify more exactly the features that change over time so that they characterise music and movement sequences: *dynamics/kinematics*, *shape/shaping*, *chunk/chunking*, *peak structure*, *pattern/patterning*, *synch-points*, *goal-points* as well as *activation* and *effort*. The review has involved music theory, the theory of non-verbal communication/linguistics, semiotics and biomechanics. Working out a basic understanding of the terms above and how they relate to each other has provided a theoretical foundation for the analytical procedures.

9.2.3. Methodology: quantitative vs a qualitative approach

The approach to analysing music and movement has been mainly qualitative. The exception is the interrater study used to evaluate consistency in correspondence judgements (chapter 6). The study demonstrated a relatively low level of interrater consistency measured with Cohen's Kappa. In addition to design issues I suggested that the low proportion of agreement may be understood as related to the way music and movement correspond in a flexible manner:

- Musical sound is characterised by a number of co-evolving features so that the same musical excerpt may afford a variety of movement responses. This implies that different criteria for correspondence may work for the same musical excerpt.
- When judging perceived correspondence I assumed that the raters were basing their judgement on perceptual *similarity* and not exact *sameness*. This adds a certain flexibility to the judgements which is not compatible with a quantification of consistency using Kappa estimation

Further methodological challenges are added when we consider that correspondence may emerge as a result of multisensory integration, i.e. that an auditory and a visual stream fuse when they are combined even though they are apparently not *similar* prior to integration. This aspect was relevant in the analysis of the free dance-movements, but not to the same extent in the sound-tracings.

The flexibility found in music-movement correspondences makes it difficult to address the consistency of such correspondences using a quantitative approach. One might have concluded that this means that music-movement correspondences are fundamentally arbitrary. However, the qualitative analyses of the sound-tracings show that there are certain features in music that correspond consistently with certain features in music. I have therefore suggested that studies of music-movement correspondences should include detailed qualitative analyses, and that this may for the present be the most reliable way of 'measuring' how music and movement work together.

9.2.4. The role of theory

In sum, the theoretical framework has been worked out in order to provide:

- a general understanding of music-movement relations, i.e. an understanding within the framework of *music perception*,
- a theoretical foundation for the concepts applied to analysis,
- and a basis for discussing the overall methodological approach to analysis

9.3. Summary of findings

The main research question was formulated in the first chapter: *How do perceived aspects of movement interact with the perceived aspects of a simultaneous, co-evolving musical sound so that the two streams are "seen-heard" as corresponding?* This has been discussed in the analyses in chapters 6, 7 and 8. The analyses have already been summarised in the concluding parts in chapters 6 and 8. Here I shall give a brief overview:

9.3.1. Changes in features that characterise music and movement

First, the discussion has built on the overall assumption that music-movement correspondences emerge on the basis of similar changes in music and movement respectively within a time-window:

- sound-tracings: i.e. by showing how sound-tracers and raters seem to respond to changes in kinematics, timbre and dynamics.
- free dance movements, where *activation contour* has been discussed as an overall feature, and where this has been further demonstrated by breaking up the overall contour into single *activation features* (such as *density of onsets*, *articulation*, *loudness* etc.)

9.3.2. The non-arbitrary

On a theoretical basis I have proposed that the *non-arbitrary* aspect of music-movement correspondences may be understood in light of the vast number of sound/movement relations we have been exposed to whether as performers or as perceivers of sound. These experiences provide a basis for a pre-understanding of music-movement correspondences. Another way of formulating this is to say that we have acquired a basic sensitivity to music-movement relations.

More specifically, I have suggested that the following features of music and movement may be understood as *non-arbitrary*:

- first, as mentioned above, matching in terms of similar changes in features within a time window
- second, and related to this, matching in terms of a recognisable contour/shape or pattern

Additionally, what one may call *amodal features*, i.e. features of a percept that are easily transferred from one sensory modality to another, seem to contribute to correspondence *non-arbitrarily*, as discussed in connection with the sound-tracings (chapter 6). However, some of the *amodal features* might be understood in terms of a change in a feature (e.g. the *amodal feature* of energy/intensity) or pattern (e.g. the *amodal features* of rhythm and number).

In the introductory chapter of the thesis I hypothesised that *points of synchronisation* would provide a special glue which fuses music and movement. Understood as moments in time at which music and movement are accentuated so that they merge into one audio-visual emphasis, the notion of synch points comprises many different interrelated features:

- Articulation, i.e. are movement and music articulated similarly (smoothly vs distinctly/sharply or degrees between)?
- Force, i.e. are movement and music at this moment weighted similarly?
- Timing, i.e. does the accent occur at the same point in time in both the music and movement?
- Pattern, i.e. the emphasis in music and movement respectively which is embedded in a *prefix – goal-point – suffix* process/pattern so that the two phenomena appear similar, e.g. that the dynamics preceding and succeeding the emphasis are similar in the visual and the auditory stream (i.e. a similarity in terms of the dynamical shape or phrase structure of which the emphasis in music and movement respectively is a part).
- Interaction effect, i.e. synch points may emerge as a result of multisensory interaction, e.g. that the timing of an accent in movement is captured by the timing of an accent in music

In sum, I would propose that these features may be viewed as *non-arbitrary* factors for correspondence but with matching in terms of similar changes in features and/or some kind of pattern/contour as a common principle.

9.3.3. The flexible

My analyses of both the sound-tracings and the free dance-movements suggest that music-movement correspondences may be perceived with a high level of flexibility, i.e. that the same musical excerpt may correspond to different variants of movement. In the previous section on methodological issues I proposed three main sources for this flexibility:

- Co-evolving features in music leading to different options in terms of movement responses.
- Interaction effects due to multisensory perception.
- The perceiver accepting *similarity* as opposed to *sameness* as a criterion for correspondence.

That co-evolving features in music may *afford* different alternatives for movement responses was illustrated in the analyses of sound-tracings. Likewise, in the analyses of free dance movements it seemed that the correspondence was related to the overall activation contour as a *non-arbitrary* aspect, but that when breaking this overall shape up into sub-shapes it seemed that the overall shape could be articulated flexibly in different ways, e.g. through changes in the *density of onsets* or through changes in *dynamics/movement character*. The audio-visual variants also demonstrated how music and movement may converge in terms of dynamics while differing in terms of kinematics.

9.4. Music and movement as perceptual reality

I have concentrated the thesis on the way music and movement correspond with respect to *non-symbolic* aspects of meaning. The theoretical foundation for the *symbolic/non-symbolic* distinction was elaborated on in chapters 4 and 5. Certainly, the *symbolic* and the *non-symbolic* will always be interwoven in any experience of music and/or human-body movement.

Nevertheless, I have argued that it has been justified to maintain a specific focus on the *non-symbolic* aspects in the empirical material of this thesis. In my view, the way I have concentrated on *non-symbolic* aspects, both theoretically and analytically, provides the thesis with a relevance that goes beyond the primary research question of the thesis. The observational studies of air-instrument playing, sound-tracings and free dance-movements were designed to facilitate an immediate movement response to the musical excerpts, and by studying spontaneous bodily responses to music we may gain insights into how music becomes meaningful to listeners, thus contributing to a general theory of music and meaning. Furthermore, since the movement response, whether imagined or real, appears to be a pertinent feature of musical experience, I believe that any attempt to work out a theory of the way music is connected to human-body movement should account for the connection in terms of *non-symbolic* aspects.

This leads to the notion that music and movement relate to each other as a perceptual reality, as proposed by Eric Clarke (Clarke, 2005). This thesis elaborates on that notion:

- by outlining a theory of the way perception and action are intimately linked in perceptual/cognitive processes,
- by focusing on the *non-symbolic* and working out analytical concepts that are applicable to both music and movement,
- and by analysing the audio-visual material with respect to similarities between music and movement.

Understanding music and movement as being linked as a *perceptual reality* is an important point which I believe is worthwhile pursuing in future discussions about the relationships between music and movement.

Appendix

Text of instructions for the sound-tracing study

Instruction text in Norwegian: ”Lyd-tegning”

”Bakgrunnen for eksperimentet er at vi tror at når vi hører en lyd så vil mange av oss forestille oss en bevegelse.

I eksperimentet vil du får høre ulike lyder som blir avspilt etter hverandre. Lyd-eksemplene er til å begynne med korte, fra et halvt sekund til et par sekunder. Etter hvert blir de lengre, med varighet opp til fem sekunder. Lydene er både umusikalske/dagligdagse lyder, og de er lyder av musikk-instrumenter. Mellom hver lyd vil det være en liten pause med varighet som tilsvarer den nettopp avspilte lyden.

Foran deg har du et brett med en lys grå flate. Når du tegner på denne med pennen vil datamaskinen registrere bevegelsen du gjør på brettet med pennespissen. Du vil ikke selv se dette på brettet.

Vi skal nå sette i gang avspilling av lyd-eksemplene, og vi ber deg i pausen etter hver lyd om å gjøre den bevegelsen på brettet som du synes passer med lyden. Vi ber deg om å tegne lyden slik du spontant oppfatter den.”

English translation of sound-tracing instructions

”The purpose of the experiment is to explore the notion that listening to sounds evokes an imagery of movement.

In the experiment you are going to listen to a series of different sounds that are played in succession. The first sound examples are short, with a duration from half-a-second to two seconds. They become longer in duration, up to five seconds. The sounds are both everyday/non-musical sounds, as well as sounds on musical instruments. Between each sound there is a short break of the same duration as the sound that was just played.

In front of you there is a tablet with a light grey surface. When you draw on the tablet with this pen the computer will record the movement you make on the tablet. The drawing you make will not be visible on the tablet.

When we start playing the sound examples, we would ask you in the short break after each sound to make the movement on the tablet that you think corresponds best to the sound. We would ask you to draw the sound the way you experience it spontaneously.”

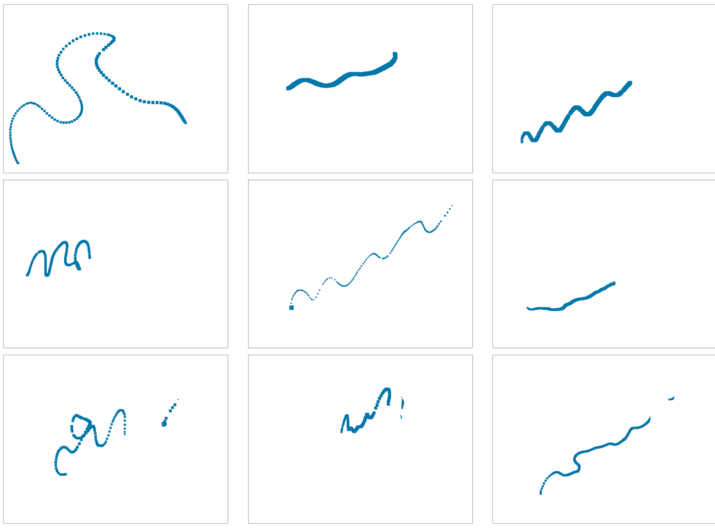
Texts of instructions for the interrater study

Instruction text in Norwegian

"På arket foran deg ser du ni forskjellige tegninger, som er tegnet av ni forskjellige personer. De ble samlet i et eksperiment der deltagerne, en og en, fikk høre en serie med lydeksempler. Lydeksemplene var korte, fra under ett sekund og opptil 5-6 sekunder. Deltagerne ble bedt om for hver lyd de hørte å tegne en bevegelse som de mente passet til lydeksempelet. Så har vi gått gjennom materialet og satt sammen de ni tegningene som hørte til en og samme lyd på ett ark som du har foran deg. Altså: på arket har du ni forskjellige tegninger som er ment å passe til det samme lydeksempelet.

I dette eksperimentet skal du få se et utvalg av slike ark med ni tegninger på hver, og du skal få høre det tilhørende lydeksempelet. Når jeg spiller av lyden ber vi deg lytte til lyden og samtidig se på tegningene på arket. Så ber vi deg vurdere hvilke 3 tegninger på arket du synes passer best til lydeksempelet. Disse merker du med tallet 1. Så ber vi deg vurdere hvilke tegninger du synes passer dårligst til lydeksempelet. Disse merker du med tallet 3. Du bestemmer selv hvor mange ganger du trenger å høre lyden. Når du har gjort dine valg blar du videre til neste eksempel.

Det første eksemplet er en prøve; lytt til lyden og se på arket."

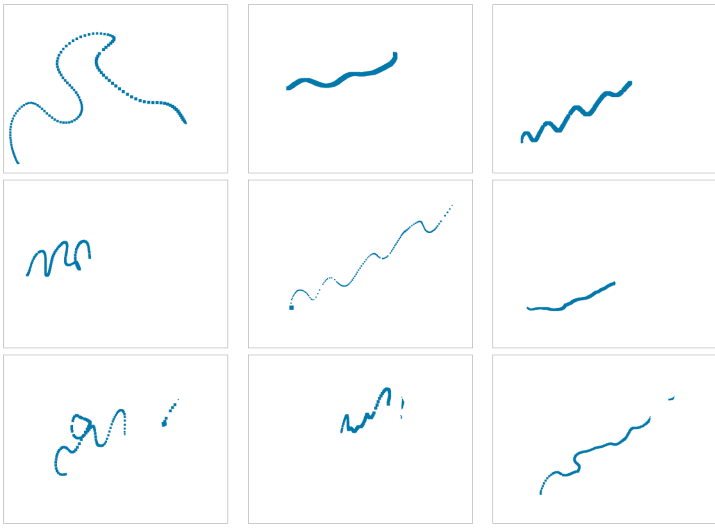


English translation of instructions for the interrater study

On the paper in front of you you can see a display of nine different drawings, each of them drawn by a different person. They were collected during an experiment in which the participants, one by one, heard a series of sound examples. The sound examples were short in duration, lasting from less than a second up to 5-6 seconds. The participants were asked for each sound to make a movement on a tablet which they thought matched the sound best. We have re-arranged the collected drawings so that nine drawings corresponding to one sound appear on the paper you have in front of you; in other words, the nine different drawings correspond to the same sound example.

In this experiment we will present a selection of such displays to you with nine drawings and the corresponding sound example. When you hear the sound we would ask you to look at the drawings and choose three drawings that you think match the sound best. We would ask you to mark these drawings with the number 1. Then we would ask you to choose three drawings that you think match the sound least. These should be marked with the number 3. You are allowed to hear the sound as many times as you wish. When you have made your choice for one display, please proceed to the next.

The first display is a test: please listen to the sound and look at the drawings."



Computation of Cohen's Kappa

Computation builds on the diagonal (bold numbers) that indicate the number of full agreement between the two raters.

		R12			Row totals
		1	2	3	
R15	1	**34**	11	6	51
	2	13	**25**	13	51
	3	4	15	**32**	51
Column totals		51	51	51	153

In this example the total number of agreements is obtained as follows:
$\Sigma a = 34 + 25 + 32 = 91$
This gives a percentage of observed agreement:
Σa/total number of ratings = 91/153 = 0,59, i.e. 59 %

Next, the expected frequency for the number of agreements that would have been expected by chance for each coding category is computed:
ef = row total x column total / overall total (N) = 51 x 51 / 153 = 17

In this example the value is the same for each coding category, so that the sum of expected frequencies of agreement by chance is given in:
$\Sigma e = 17 + 17 + 17 = 51$

Cohen's Kappa is computed as follows:
$K = \Sigma a - \Sigma e / N - \Sigma e = 91 - 51 / 153 - 51 = 40 / 102 = 0,392$

References

Adorno, T., & Eisler, H. (1994). *Composing for the Films*. London: Athlone press.

Atkinson, R. L., Atkinson, R. C., & Hilgard, E. R. (1983). *Introduction to Psychology*. San Diego London: Harcourt Brace Jovanovich, Publishers.

Bengtsson, I., & Gabrielsson, A. (1983). Analysis and Synthesis of Musical Rhythm. In J. Sundberg (Ed.), *Studies of Music Performance* (Vol. No 39, pp. 27 - 60). Stockholm: Publications issued by the Royal Swedish Academy of Music.

Berthoz, A. (2000). *The Brain's Sense of Movement* (G. Weiss, Trans.). Cambridge, MA: Harvard University Press.

Bower, T. (1974). *Development in Infancy*. San Francisco: Freeman & Co.

Bowie, A. (2002). Music and the Rise of Aesthetics. In J. Samson (Ed.), *The Cambridge History of Nineteenth-Century Music* (pp. 29-54). Cambridge: Cambridge University Press.

Branigan, E. (1997). Sound, Epistemology, Film. In R. Allen & M. Smith (Eds.), *Film Theory and Philosophy* (pp. 95 - 125). Oxford: Clarendon Press.

Bregman. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Cadoz, C., & Wanderley, M. (Eds.). (2000). *Gesture-Music*. Paris: IRCAM-Centre Pompidou.

Calvert, G. A. (2001). Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. *Cerebral Cortex, 11*, 1110-1123.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response Amplification in Sensory-Specific Cortices During Crossmodal Binding. *Neuroreport, 10*, 2619-2623.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal Identification. *Trends in cognitive science, 2*(7), 247 -253.

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The Handbook of Multisensory Processes*. Cambridge, MA London: A Bradford Book The MIT Press.

Carello, C., Wagman, J., & Turvey, M. T. (2005). Acoustic Specification of Object Properties. In J. Anderson & B. Anderson (Eds.), *Moving Image Theory: Ecological Considerations* (pp. 79-104). Carbondale: University of Southern Illinois Press.

Casciato, C. (2005). Towards an Understanding of Gesture and Timbre: Mc Gill University, Music Technology.

Chion, M. (1994). *Audio-Vision. Sound on Screen* (C. Gorbman, Trans.). New York: Columbia University Press.

Clarke, E. (1999). Rhythm and Timing in Music. In Deutsch (Ed.), *The Psychology of Music* (2nd ed.). San Diego: Academic Press.

Clarke, E., & Davidson, J. (1998). The Body in Performance. In W. Thomas (Ed.), *Composition, Performance, Reception* (pp. 74-92). Aldershot: Ashgate.

Clarke, E. F. (1993). Generativity, Mimesis and the Human Body in Music Performance. *Contemporary music review, vol 9*(part 1 & 2), 207-218.

Clarke, E. F. (2005). *Ways of Listening. An Ecological Approach to the Perception of Musical Meaning*. New York: Oxford University Press.

Clayton, M., Sager, R., & Will, U. (2004). In Time with the Music: The Concept of Entrainment and Its Significance for Ethnomusicology. *ESEM CounterPoint, 1*.

Coker, W. (1972). *Music and Meaning*.

Cook, N. (1998). *Analysing Musical Multimedia*. Oxford: Oxford University Press.

Danielsen, A. (2006). *Presence and Pleasure: The Funk Grooves of James Brown and Parliament*. Middleton, Conn.: Wesleyan University Press.

Davidson, J. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. *Psychology of music, 21*, 103-113.

Deutsch, D. (1999). *The Psychology of Music* (2nd ed.). San Diego: Academic Press.

Dowling, W. J. (1994). Melodic Contour in Hearing and Remembering Melodies. In J. A. Sloboda & R. Aiello (Eds.), *Musical Perceptions* (pp. 173-190). New York: Oxford University Press.

Eichert, R., Schmidt, L., & Seifert, U. (1997). Logic, Gestalt Theory, and Neural Computation in Research on Auditory Perceptual Organization. In M. Leman (Ed.), *Music, Gestalt and Computing* (pp. 70-88). Heidelberg: Springer-Verlag.

Eisenstein, S. (1986). *The Film Sense*. London   Boston: Faber and Faber.

Eitan, Z., & Granot, R. Y. (2006). How Music Moves: Musical Parameters and Listener's Images of Motion. *Music perception, 23*(3), 221-247.

Feldman, J., Epstein, D., & Richards, W. (1992). Force Dynamics of Tempo Change in Music. *Music perception, vol 10*(no 2), 185-203.

Fendrich, R., & Corballis, P. M. (2001). The Temporal Cross-Capture of Audition and Vision. *Perception & Psychophysics, 63*(4), 719-725.

Frazier, L., Carlson, K., & Clifton Jr, C. (2006). Prosodic Phrasing Is Central to Language Comprehension. *Trends in cognitive science 10* (6). 244-249

Gabrielsson, A. (1986). Rhythm in Music. In J. Evans & M. Clynes (Eds.), *Rhythm in Psychological and Musical Processes* (pp. 131 - 166). Springfield: C.C. Thomas.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Godøy, R. I. (1997). Knowledge in Music Theory by Shapes of Musical Objects and Sound-Producing Actions. In M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 89-102). Heidelberg: Springer-Verlag.

Godøy, R. I. (1999). Shapes and Spaces in Musical Thinking: University of Oslo, Department of Musicology. Unpublished

Godøy, R. I. (2001). Imagined Action, Excitation and Resonance. In R. I. Godøy & H. Jørgensen (Eds.), *Musical Imagery* (pp. 237-250). Lisse (Holland): Swets & Zeitlinger.

Godøy, R. I. (2006). Gestural-Sonorous Objects: Embodied Extensions of Schaeffer's Conceptual Apparatus. *Organised Sound, 11*(2), 149-157.

Godøy, R. I. (2008a). Gestural Affordances of Musical Sound. In R. I. Godøy & M. Leman (Eds.), *Sound Gestures*.

Godøy, R. I. (2008b). Reflections on Chunking. In A. Schneider (Ed.), *Hamburger Jahrbuch Für Musikwissenschaft 24*. Frankfurt: Peter Lang.

Godøy, R. I., Haga, E., & Jensenius, A. R. (2006a). Exploring Music-Related Gestures by Sound-Tracing. A Preliminary Study. In K. Ng (Ed.), *Proceedings of the Cost-287 Congas 2nd International Symposium on Gesture Interfaces for Multimedia Systems Gims* (pp. 27-33). Leeds.

Godøy, R. I., Haga, E., & Jensenius, A. R. (2006b). Playing "Air Instruments": Mimicry of Sound-Producing Gestures by Novices and Experts. In S. Gibet, N. Courty & J.-F. Kamp (Eds.), *6th International Gesture Workshop. Gesture in Human-Computer Interaction and Simulation. Berder Island, France, May 18-20, 2005* (pp. 256-267). Heidelberg: Springer-Verlag.

Hackney, P. (2002). *Making Connections. Total Body Integration through Bartenieff Fundamentals*. New York London: Routledge.

Harnad, S. (1987). Category Induction and Representation. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 535-565). Cambridge: Cambridge University Press.

Hatten, R. S. (2004). *Interpreting Musical Gestures, Topics, and Tropes. Mozart, Beethoven, Schubert*. Bloomington and Indianapolis: Indiana University Press.

Heft, H. (2001). *Ecological Theory in Context: James Gibson, Roger Barker, and the Legacy of William James' Radical Empiricism*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Iwamiya, S.-i. (1994). Interactions between Auditory and Visual Processing When Listening to Music in Audio Visual Context: 1. Matching 2. Audio Quality. *Psychomusicology, 13*, 133-154.

Jensenius, A. R. (2006). Using Motiongram in the Study of Musical Gestures., *International Computer Music Conference*. New Orleans.

Jensenius, A. R. (2008). *Action-Sound. Developing Methods and Tools for Studying Music-Related Body Movement.* Unpublished PhD, University of Oslo, Oslo.

Jensenius, A. R., Godøy, R. I., & Wanderley, M. M. (2005). *Developing Tools for Studying Musical Gestures within the Max/Msp/Jitter Environment.* Paper presented at the The international computer music conference, 4-10 September, Barcelona.

Jeppesen, K. (1992). *Counterpoint. The Polyphonic Vocal Style of the Sixteenth Century*. New York: Dover Publications.

Johansson, G. (1973). Visual Perception of Biological Motion and a Model for Its Analysis. *Perception & Psychophysics, Vol 14, No 2*, 201-211.

Johnson, M. (1987). *The Body in the Mind.* Chicago: University of Chicago Press.

Kamitani, Y., & Shimojo, S. (2001). Sound-Induced Visual "Rabbit". *Journal of vision, 1*(3), 478a.

Kendon, A. (1996). An Agenda for Gesture Studies. *The semiotic review of books, 7*(3), 7-12.

Kendon, A. (2004). *Gesture. Visible Action as Utterance*. Cambridge: University Press.

Kipp, M. (2004). *Gesture Generation by Imitation - from Human Behavior to Computer Character Animation.* Unpublished Doctoral, Saarland University, Saarbrücken, Germany.

Kitagawa, N., & Ichihara, S. (2002). Hearing Visual Motion in Depth. *Nature, 416*, 172-174.

Kronman, U., & Sundberg, J. (1987). *Is the Musical Ritard an Allusion to Physical Motion?* Paper presented at the Symposium in the Third International Conference on Event Perception and Action. Action and Perception in Rhythm and Meter. Stockholm: Royal Swedish Academy of Music, 1987, No. 55, 57-68.

Krumhansl, C. L. (1996). A Perceptual Analysis of Mozart's Piano Sonata K 282: Segmentation, Tension and Musical Ideas. *Music perception, 13*(3), 401-432.

Krumhansl, C. L., & Schenk, D. L. (1997). Can Dance Reflect the Structural and Expressive Qualities of Music? A Perceptual Experiment on Balanchine's Choreography of Mozart's Divertimento No 15. *Musicae Scientiae 1* (1), 63-85.

Laban, R. (1948). *Modern Educational Dance*. London: MacDonald&Evans.

Laban, R. (1971). *The Mastery of Movement*. London: MacDonald & Evans.

Laban, R., & Lawrence, F. (1947). *Effort*. London: MacDonald & Evans.

Laban, R., & Ullmann, L. (1966). *Choreutics*. London: MacDonald & Evans.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1984). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Leman, M. (1993). Symbolic and Subsymbolic Description of Music. In G. Haus (Ed.), *Music Processing* (pp. 119-164). Oxford: Oxford University Press.

Lewkowicz, D. J., & Kraebel, K. S. (2004). The Value of Multisensory Redundancy in the Development of Intersensory Perception. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 655-678). Cambridge, MA London: A Bradford Book The MIT Press.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-Modal Equivalence in Early Infancy: Audio-Visual Intensity Matching. *Developmental psychology, 16*(597-607).

Liberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception Revised. *Cognition, 21*, 1-36.

Lickliter, R., & Bahrick, L. E. (2004). Perceptual Development and the Origins of Multisensory Responsiveness. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 643-653). Cambridge, MA London: A Bradford Book The MIT Press.

Lipscomb, S. D., & Kim, E. M. (2004). *Perceived Match between Visual Parameters and Auditory Correlates: An Experimental Multimedia Investigation.* Paper presented at the The 8th International Conference on Music Perception and Cognition, Northwestern University.

Manabe, K., & Riquimaroux, H. (2000). Sound Controls Velocity Percpetion of Visual Apparent Motion. *Journal of acoustical society of Japan, 21*(3), 171-174.

Marshall, S. K., & Cohen, A. J. (1988). Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures. *Music perception, 6*(1), 95-112.

Mateef, S., Hohnsbein, J., & Noack, T. (1985). Dynamic Visual Capture: Apparent Auditory Motion Induced by a Moving Visual Target. *Perception, 14*, 721-727.
Mathews, M. (2001). What Is Loudness. In P. R. Cook (Ed.), *Music, Cognition and Computerized Sound* (pp. 71-78). Cambridge MA: MIT Press.
McAdams, S. (1984). *Spectral Fusion, Spectral Parsing, and the Formation of Auditory Images.* Stanford university.
McAdams, S., Depalle, P., & Clarke, E. (2004). Analyzing Musical Sound. In E. Clarke & N. Cook (Eds.), *Empirical Musicololgy. Aims, Methods, Prospects* (pp. 157-196). Oxford: Oxford University Press.
McGurk, H., & MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature, 264*, 746-748.
McNeill, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chigago: The University of Chicago Press.
Meltzoff, A. N., & Borton, W. (1979). Intermodal Matching by Human Neonates. *Nature, 282*, 403-404.
Merleau-Ponty, M. (1983). *The Structure of Behavior* (A. L. Fisher, Trans.). Pittsburgh, PA: Duquesne University Press.
Metz, C. (1985). Aural Objects. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice*. New York: Columbia University Press.
Meyers, R. (1994). *Film Music. Fundamentals of the Language*. Oslo: Ad notam Gyldendal.
Mikumo, M. (1998). Encoding Strategies for Pitch Information. *Japanese psychological monographs*(No 27).
Mitchell, R. W., & Gallaher, M. C. (2001). Embodying Music: Matching Music and Dance in Memory. *Music perception, 19*(1), 65-85.
Moore, C.-L., & Yamamoto, K. (1988). *Beyond Words*. New York London: Routledge.
Noë, A. (2004). *Action in Perception*. Cambridge MA London: MIT Press.
Pavlicevic, M. (1990). Dynamic Interplay in Clinical Improvisation. *British Journal of Music Therapy, 4*(2), 5-9.
Pedhazur, E., & Schmelkin, L. (1991). *Measurement, Design, and Analysis. An Integrated Approach*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
Pollick, F. (2004). The Features People Use to Recognise Human Movement Style. In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction* (pp. 10-19). Heidelberg: Springer-Verlag.
Rasch, R., & Plomp, R. (1999). The Perception of Musical Tones. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 89-112). San Diego: Academic Press.
Redfern, B. (1965). *Introducing Laban Art of Movement*. London: MacDonald & Evans.
Repp, B. (1998). A Microcosm of Musical Expression. I. Quanitative Analysis of Pianists' Timing in the Initial Measures of Chopin's Etude in E Major. *Journal of the acoustical society of America, 104*(2).

Repp, B. (1999). A Microcosm of Musical Expression: Ii. Quantitative Analysis of Pianist's Dynamics in the Initial Measures of Chopin's Etude in E Major. *Journal of Acoustical society of America, 105*(3), 1972 - 1988.

Rimsky-Korsakoff, N. (1964). *Principles of Orchestration* (E. Agate, Trans.). New York: Dover Publications, Inc.

Rizzolatti, G., Gartilucci, M., Camarda, R. M., Gallex, G., Luppino, G., Matelli, M., et al. (1990). Neurons Related to Reaching-Grasping Arm Movements in the Rostral Part of Area 6 (Are 6a). *Experimental brain research, 82*, 337-350.

Roads, C. (1996). *The Computer Music Tutorial*. Cambridge, MA: MIT Press.

Rosenbaum, D. A. (1991). *Human Motor Control*. San Diego: Academic Press.

Runeson, S., & Frykholm, G. (1983). Kinematic Specification of Dynamics as an Informational Basis for Person-and Action Perception: Expectation, Gender Recognition and Deceptive Intention. In. *Journal of experimental psychology, General, vol 112, no 4*, 585 - 615.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound Alters Visual Motion Perceptioin. *Nature, 385*, 308.

Shaffer, L. (1981). Performances of Chopin, Bach and Bartok: Studies in Motor Programming. *Cognitive psychology, 13*, 326-376.

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of Visual Perception by Sound. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 27-33). Cambridge, MA London: A Bradford Book The MIT Press.

Shipley, T. (2003). The Effect of Object and Event Orientation on Perception of Biological Motion. *Psychological science, 14*, 377 - 388.

Shove, P., & Repp, B. (1995). Musical Motion and Performance: Theoretical and Empirical Perspectives. In Rink (Ed.), *The Practice of Performance* (pp. 55-83). Cambridge: Cambridge University Press.

Sirius, G., & Clarke, E. F. (1994). The Perception of Audiovisual Relationships: A Preliminary Study. *Psychomusicology, 13*, 119-132.

Snyder, B. (2000). *Music and Memory*. Cambridge: The MIT Press.

Soto-Faraco, S., & Kingstone, A. (2004). Multisensory Integration of Dynamic Information. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 49-67). Cambridge, MA London: A Bradford Book The MIT Press.

Stein, B., & Meredith, M. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of cognitive neuroscience, 8*(6), 497-506.

Stern, D. (2000). *The Interpersonal World of the Infant*. New York: Basic books.

Stern, D. (2004). *The Present Moment. In Psychotherapy and Everyday Life*. New York: W. W. Norton & Company, Inc.

Sundberg, J., & Verillo. (1980). On the Anatomy of the Retard: A Study of Timing in Music. *Journal of the acoustical society of America, 68*(3), 772-779.

Todd, N. P. McA. (1992). The Dynamics of Dynamics: A Model of Musical Expression. *Journal of the acoustical society of America, 91(6)*, 3540-3550.

Todd, N. P. McA. (1995). The Kinematics of Musical Expression. *Journal of the acoustical society of America, 97(3)*, 1940-1949.
Trondalen, G., & Skårderud, F. (2007). Playing with Affects. And the Importance of "Affect Attunement". *Nordic Journal of Music Therapy, 16*(2), 100-111.
Uebersax, J. (2006). Statistical Methods for Rater Agreement. from http://ourworld.compuserve.com/homepages/jsuebersax/agree.htm
Varela, F. J., Thompson, E., & Rosch, E. (1993). *The Embodied Mind*. Cambridge, MA: MIT Press.
Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2005). Cross-Modal Interactions in the Perception of Musical Performance. *Cognition, 101*, 80-113.
Vroomen, J., & de Gelder, B. (2000). Sound Enhances Visual Perception: Cross-Modal Effects of Auditory Organization on Vision. *Journal of experimental psychology, Human Perception and Performance, Vol 26, No 5*, 1583-1590.
Wagner, R. (1979). Music of the Future (R. Jacobs, Trans.). In *Three Wagner Essays*. London: Eulenburg Books.
Wallace, M. T. (2004). The Development of Multisensory Integration. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 625-642). Cambridge, MA London: A Bradford Book The MIT Press.
Welch, R. B., & Duttonhurt, L. D. (1986). Contributions of Audition and Vision to Temporal Rate Perception. *Perception & Psychophysics, 39*(4), 294-300.
Wilson, M., & Knöblich, G. (2005). The Case for Motor Involvement in Perceiving Conspecifics. *Psychological bulletin, 131*(3), 460-473.
Winter, D. A. (2005). *Biomechanics and Motor Control of Human Movement* (3rd ed.). New York: John Wiley & Sons.
Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action Generation and Action Perception in Imitation: An Instance of the Ideomotor Principle. *The Royal Society (online)*.
Waadeland, C. (2000). *Rhythmic Movements and Moveable Rhythms. Syntheses of Expressive Timing by Means of Rhtymic Frequency Modulation.* Unpublished Dr Art, Norwegian University of Science and Technology, Trondheim.